

Mammalian Grainyhead transcription factors

**FIELD OF THE INVENTION**

- 5 The present invention relates generally to diagnostic and therapeutic agents. More particularly, the present invention provides mammalian transcription factors which function in the modulation of expression of genetic sequences. The present invention further provides nucleic acid molecules encoding the transcription factors as well as nucleic acid and/or proteinaceous molecules with which the transcription factors interact.
- 10 The transcription factors of the present invention or molecules interacting with same may be used *inter alia* in the generation of a range of diagnostic and therapeutic agents for a range of conditions. Therapeutic agents include gene-expression modulating agents including sense and antisense molecules, ribozymes and RNAi - type molecules. The present invention further provides medical assessment systems including drug evaluation
- 15 systems comprising genetically modified animals.

**BACKGROUND OF THE INVENTION**

- Reference to any prior art in this specification is not, and should not be taken as, an
- 20 acknowledgment or any form of suggestion that this prior art forms part of the common general knowledge in any country.

Bibliographic details of references provided in the subject specification are listed at the end of the specification.

25

- The increasing sophistication of recombinant DNA techniques has provided significant progress in understanding the mechanisms involved in regulating eukaryotic gene expression. This is greatly facilitating research and development in the plant, agricultural, medical and veterinary industries. Transcription factors are an important component in the
- 30 control of gene expression. However, despite their importance, mammalian transcription factors have not been well investigated for their diagnostic and therapeutic potential.

- 2 -

RNA polymerases in eukaryotic cells cannot initiate transcription alone; before transcription can begin, they require interaction between transcription factors and the promoter. These factors assemble at the promoter and, *via* a series of steps, facilitate both  
5 the binding of RNA polymerase II to the promoter and its subsequent phosphorylation and release to initiate transcription.

In addition to these general transcription factors, many thousands of transcription activators and/or negative regulators (inhibitors) exist, which control the process of  
10 initiation of gene transcription from great distances along the DNA. These factors influence the timing and extent of transcription of a particular gene. Indeed, they control whether and to what extent a particular gene is transcribed in a cell of a particular tissue type. Although most gene regulators identified to date have been found to be proteins, some transcription factors may also be RNA molecules.

15

In *Drosophila*, the transcription factor known as "Grainyhead" regulates key developmental process in the embryo and is encoded by the gene *grainyhead*. During development, Grainyhead is initially involved in dorsal/ventral and terminal patterning of the newly fertilized embryo through the formation of multi-protein complexes that repress  
20 transcription from the *decapentaplegic*, *tailless* and *zerknuehlt* genes (Huang *et al.*, *Genes Dev.* 9: 3177-3189, 1995; Liaw *et al.*, *Genes Dev.* 9: 3163-3176, 1995). Later, *grainyhead* is predominantly expressed in the embryonic central nervous system in cuticle-producing tissues, where it binds to promoters and influences transcription from other developmentally regulated genes including *engrailed*, *fushi tarazu* and *Ultrabithorax* (Bray  
25 *et al.*, *Genes Dev.* 3: 1130-1145, 1989; Dynlacht *et al.*, *Genes Dev.* 3: 1677-1688, 1989; Biggin and Tjian, *Cell* 53: 699-711, 1988; Soeller *et al.*, *Genes Dev.* 2: 68-81, 1988; Dynlacht *et al.*, *Cell* 56: 563-576, 1991; Attardi and Tjian, *Genes Dev.* 7: 1341-1353, 1993; Uv *et al.*, *Mol. Cell Biol.* 14: 4020-4031, 1994).

- 3 -

The importance of *grainyhead* in *Drosophila* development is emphasised by the embryonic lethal phenotype observed in flies carrying mutations in this gene. The embryos have flimsy cuticles, grainy and discontinuous head skeletons and patchy tracheal tubes (Bray and Kafatos, *Genes Dev.* 5: 1672-1683, 1991). A neuroblast-specific isoform of the protein, arising from alternate splicing, has also been identified. A mutation that abolishes this isoform is pupal- and adult- lethal, and flies demonstrate uncoordinated movements (Uv *et al.*, *Mol. Cell Biol.* 17: 6727-6735, 1997).

Mammalian homologs of *grainyhead* have previously been proposed, including three genes designated *CP2*, *LBP-1a* and *LBP-9*. Studies have implicated them in a wide variety of cellular and developmental events including T cell proliferation, globin gene expression and steroid biosynthesis (Sueyoshi *et al.*, *Mol. Cell Biol.* 15: 4158-4166, 1995; Jane *et al.*, *EMBO J.* 14: 97-105, 1995; Volker *et al.*, *Genes Development* 11: 1435-1446, 1997; Zhou *et al.*, *Mol. Cell Biol.* 20: 7662-7672, 2000). However, *in situ* analyses of both *CP2* and *LBP-1a* reveal ubiquitous expression of both genes, unlike the highly restricted pattern observed with *grainyhead* in *Drosophila* (Bray *et al.*, 1989, *supra*; Dynlacht *et al.*, 1989, *supra*; Bray and Kafatos, 1991, *supra*; Ramamurthy *et al.*, *J. Biol. Chem.* 276: 7836-7842, 2001). It is concluded, therefore, that these genes are not close homologs of *grainyhead*.

Abnormalities in mammalian transcription factor expression are considered to play a role in a number of different genetic disorders and birth defects such as spina bifida and anencephaly. There is therefore a need to identify mammalian transcription factors and in particular close mammalian homologs of Grainyhead and to use these to develop a range of diagnostic and therapeutic agents.

- 4 -

## SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word “comprise”, or variations such as “comprises” or “comprising”, will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Nucleotide and amino acid sequences are referred to by a sequence identifier number (SEQ ID NO:). The SEQ ID NOs: correspond numerically to the sequence identifiers <400>1 (SEQ ID NO:1), <400>2 (SEQ ID NO:2), etc. A sequence listing is provided at the end of the specification. A summary of the SEQ ID NOs is provided in Table 1.

Genetic sequences were studied which exhibited homology at the nucleotide and/or amino acid level to a *Drosophila* gene, the product of which, is involved in body patterning where a fine balance between activation and inhibition of gene expression is critical to the correct development of cells and tissues into functional organisms. A large number of different families of transcription factors play a critical role in ensuring that this balance is maintained during embryological development. One such transcription factor, cloned from *Drosophila* and well-characterized, is Grainyhead (hereinafter referred to by its abbreviation, GRH). GRH is encoded by the gene *grainyhead* (*grh*). The inventors observed that the identity of previously published putative *grh* mammalian homologs showed much more ubiquitous expression compared with the highly restricted pattern exhibited by *Drosophila grh*. Furthermore, sequence similarity between the proposed mammalian homologs and the *Drosophila grh* sequence was relatively low. In accordance with the present invention, true *grh* homologs were identified and derived from mammalian tissue such as human and mouse tissue.

Accordingly, one aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a mammalian homolog of *Drosophila* GRH. A mammalian homolog of GRH is referred to herein as M-GRH. The corresponding gene is referred to as M-*grh*. A M-*grh* is



- 5 -

deemed a homolog of *Drosophila grh* (*D-grh*). If it comprises a nucleotide sequence having 60% or greater similarity to the nucleotide sequence of *D-grh* after optimal alignment. Likewise, a M-GRH is so defined if it comprises an amino acid sequence having 60% or greater similarity to the amino acid sequence of *Drosophila* GDH (*D-GRH*). There are four isoforms of *Drosophila grh* designated *D-grh* P1, *D-grh* P2, *D-grh* P3 and *D-grh* P4. The nucleotide sequence encoding *D-grh* is set forth in SEQ ID NO:17 and SEQ ID NO:34, SEQ ID NO:36 and SEQ ID NO:38, respectively. Mammalian sequences encompassed by the present invention include those derived from tissues of mouse and human including, for example, mouse embryo, human fetal brain and placenta, and mouse and human kidney. Reference herein to *Drosophila grh* includes any or all of its isoforms P1-P4.

The mammalian sequences identified by the present inventors show higher percentages of similarity to the *D-grh* sequence than the already identified mammalian sequences designated *CP2*, *LBP-1a* and *LBP-9*. In accordance with the present invention, it is proposed that the M-*grh* homologs disclosed are "true" *grh* homologs relative to *CP2*, *LBP-1a* and *LBP-9*. As a result of the analysis herein described, it is shown that the earlier sequences align phylogenetically with another distinct *Drosophila* factor, designated *Drosophila CP2*. A new family of transcription factors, highly conserved from *Drosophila* to human and having distinct tissue-specificity profiles, is now described in accordance with the present invention.

The true M-*grh* homologs of the present invention include mammalian *grainyhead* (gene: *mgr*; expression product: MGR), *brother of mgr* (gene: *bom*; expression product: BOM) and *sister of mgr* (gene *som*; protein: SOM). MGR has multiple isoforms including MGR p49 and MGR p70 in humans and MGR p61 in mice. A summary of the SEQ ID NOs for the M-*grh* and M-GRH molecules of the present invention are shown in Table 2. The sequences are provided in the Sequence Listing. The gene *som* and its product SOM are also referred to herein as *grhl3* and GRHL3, respectively.

- 6 -

The present invention provides, therefore, expression products of the *M-grh* genes, *mgr*, *bom* and *som* as well as derivatives and homologs thereof. This aspect of the present invention does not extend to *CP2*, *LBP-1a* or *LBP-9*.

- 5 Accordingly, another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding a polypeptide comprising a predicted amino acid sequence substantially as set forth in SEQ ID NO:2 (human MGR p49), SEQ ID NO:4 (human MGR p70), SEQ ID NO:6 (human BOM), SEQ ID NO:8 (human SOM), SEQ ID NO:10 (murine MGR p49), SEQ ID NO:12 (murine MGR p70),  
10 SEQ ID NO:14 (murine BOM) or SEQ ID NO:16 (murine SOM) or an amino acid sequence having at least about 60% similarity to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16 after optimal alignment.
- 15 The preferred nucleic acid molecules comprise sequences of nucleotides substantially as set forth in SEQ ID NO:1 (human *mgr* p49), SEQ ID NO:3 (human *mgr* p70), SEQ ID NO:5 (human *bom*), SEQ ID NO:7 (human *som*), SEQ ID NO:9 (murine *mgr* p61), SEQ ID NO:11 (murine *mgr* p70), SEQ ID NO:13 (murine *bom*) or SEQ ID NO:15 (murine *som*) or complementary forms thereof, or a nucleotide sequence having at least about 60%  
20 similarity to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 after optimal alignment or their complementary forms or a nucleotide sequence capable of hybridizing to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or complementary forms thereof under low stringency  
25 conditions. Again, this aspect of the present invention does not extend to nucleic acid molecules encoding CP2, LBP-1 and LBP-9.

The present invention further extends to recombinant forms of the M-GRH molecules. Preferred recombinant M-GRH molecules having amino acid sequences defined in  
30 parenthesis include human MGR p49 (SEQ ID NO:2), human MGR p70 (SEQ ID NO:4), human BOM (SEQ ID NO:6), human SOM (SEQ ID NO:8), murine MGR p61 (SEQ ID

- 7 -

NO:10), murine MGR p70 (SEQ ID NO:12), murine BOM (SEQ ID NO:14) and murine SOM (SEQ ID NO:16).

Reference to "M-GRH" molecules include derivatives, homologs and analogs thereof.

5

The mammalian transcription factors of the present invention are proposed to be involved in the regulation of expression of a range of genes such as but not limited to developmentally regulated genes involved in determining patterning. Some of the genes regulated encode critical products, the absence or malfunctioning of which, is proposed to lead to unwanted phenotypes and/or predispositions to certain medical conditions. That is, the presence of a mutation in and/or malfunction of a *M-grh* including over or under expression of the transcription factors of the present invention are proposed to cause incorrect regulation of one or more of these genes thereby leading to an inappropriate phenotype. The ability to detect mutations in the nucleotide sequences encoding the *M-grh* homologs permits the detection of a range of abnormalities or a predisposition for development of abnormalities. Furthermore, as many of the genes will be developmentally regulated genes, identification of the transcription factors permits identification of unknown developmentally regulated genes.

20 Accordingly, another aspect of the present invention contemplates a method for detecting a variation in a polynucleotide sequence encoding a M-GRH transcription factor.

Furthermore, the isolated nucleic acid molecules of the present invention may be able to be used to correct such an abnormality in a subject in need thereof or at risk of developing an abnormality. The nucleic acid molecules of the present invention may be comprised, therefore, within a suitable vector for delivery of all or part of the sequence to a recipient cell or tissue. The nucleic acid molecule or part thereof could also be administered directly for transient expression. The present invention provides, therefore, the potential for both a diagnostic and a therapeutic capability.

30

- 8 -

Accordingly, a further aspect of the present invention contemplates a genetic construct comprising a nucleotide sequence selected from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15 or a nucleotide sequence having at least 60% similarity to one or more of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 after optimal alignment or a nucleotide sequence capable of hybridizing to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or a complementary form thereof under low stringency conditions.

In a related embodiment, the present invention provides a genetic construct comprising a promoter or functional equivalent thereof operably linked to a nucleotide sequence of the invention.

The present invention further provides animal models comprising genetically altered *M-grh* sequences including insertions, deletions, additions, and substitutions. Such animal models including genetically modified animals are useful in the development of medical assessment systems such as to monitor physiological changes and to evaluate drug targets and drug candidates. The medical assessment system may also be used in drug development.

Examples of drugs or other therapeutic agents include genetic agents such as sense and antisense molecules, ribozymes, DNazymes, methylakion- or demethylation- inducing agents as well as RNAi-type agents. Peptide mimetics and non-proteinaceous chemical entities are also contemplated by the present invention.

Genes are represented herein in lower case italics. Expression products (e.g. proteins or RNA) are represented in upper case, non-italic letters. A summary of the genes and their expression products is provided in Table 1. The gene "*som*" or its expression product "SOM", are also referred to as *grhl3* and GRHL3, respectively.

**TABLE 1**  
**Abbreviations**

GENE	EXPRESSION PRODUCT
<i>grainyhead</i> ( <i>grh</i> )	Grainyhead (GRH)
mammalian <i>grainyhead</i> homologs (M- <i>grh</i> )	mammalian grainyhead homologs (M-GRH)
mammalian <i>grainyhead</i> ( <i>mgr</i> )	mammalian Grainyhead (MGR)
brother of mammalian <i>grainyhead</i> ( <i>bom</i> )	brother of mammalian grainyhead (BOM)
sister of mammalian <i>grainyhead</i> ( <i>som</i> )	sister of mammalian grainyhead (SOM)

- 5 A summary of sequence identifiers used throughout the specification is Table 2.

**TABLE 2**  
**Summary of sequence identifiers**

SEQUENCE ID NO:	NAME	DESCRIPTION
1	human <i>mgr</i> p49	Nucleotide sequence encoding mammalian <i>grainyhead</i> derived from human fetal brain
2	human MGR p49	Predicted amino acid sequence corresponding to SEQ ID NO:1
3	human <i>mgr</i> p70	Nucleotide sequence encoding mammalian <i>grainyhead</i> being an isoform of SEQ ID NO:1, derived from human kidney
4	human MGR p70	Predicted amino acid sequence corresponding to SEQ ID NO:3
5	human <i>bom</i>	Nucleotide sequence encoding mammalian <i>grainyhead</i> derived from human placenta
6	human BOM	Predicted amino acid sequence corresponding to SEQ ID NO:5
7	human <i>som</i>	Nucleotide sequence encoding mammalian <i>grainyhead</i>
8	human SOM	Predicted amino acid sequence corresponding to SEQ ID NO:7
9	murine <i>mgr</i> p61	Nucleotide sequence encoding mammalian <i>grainyhead</i> derived from 17.5 day murine embryo
10	murine MGR p61	Predicted amino acid sequence corresponding to SEQ ID NO:9

- 10 -

SEQUENCE ID NO:	NAME	DESCRIPTION
11	murine <i>mgr</i> p70	Nucleotide sequence encoding mammalian <i>grainyhead</i> being an isoform of SEQ ID NO:9, derived from murine kidney
12	murine MGR p70	Predicted amino acid sequence corresponding to SEQ ID NO:11
13	murine <i>bom</i>	Nucleotide sequence encoding mammalian <i>grainyhead</i> derived from a murine embryonic carcinoma cell line (p19)
14	murine BOM	Predicted amino acid sequence corresponding to SEQ ID NO:13
15	murine <i>som</i>	Nucleotide sequence encoding mammalian <i>grainyhead</i> .
16	murine SOM	Predicted amino acid sequence corresponding to SEQ ID NO:15
17	<i>grh</i> -P1	Nucleotide sequence encoding the <i>Drosophila</i> transcription factor designated <i>Grainyhead</i> ( <i>grh</i> )
18	GRH-P1	Amino acid sequence corresponding to SEQ ID NO:18
19-20	human p49 <i>mgr</i>	primers
21-22	human p70 <i>mgr</i>	primers
23-24	human <i>bom</i>	primers
25-26	murine p70 <i>mgr</i>	primers
27-28	murine p61 <i>mgr</i>	primers
29-30	murine <i>bom</i>	primers
31-32	human S14	primers
33	<i>Drosophila dopa decarboxylase</i>	promoter
34	<i>Drosophila PCNA</i>	promoter
35	human Engrailed-1	promoter
36	<i>grh</i> -P2	Nucleotide sequence encoding the <i>Drosophila</i> transcription factor designated <i>Grainyhead</i> ( <i>grh</i> ) isoform P2
37	GRH-P2	Amino acid sequence corresponding to SEQ ID NO:36
38	<i>grh</i> -P3	Nucleotide sequence encoding the <i>Drosophila</i> transcription factor designated <i>Grainyhead</i> ( <i>grh</i> ) isoform P3
39	GRH-P3	Amino acid sequence corresponding to SEQ ID NO:38

- 11 -

SEQUENCE ID NO:	NAME	DESCRIPTION
40	GRHL-3	Primer
41	GRHL-3	Primer
42	GRHL-3	Primer
43	HPRT	Primer
44	Antisense	Primer
45	Exon 8 and Exon 13 Sense	Primer
46	Antisense	Primer

## BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation showing that *mgr* genomic locus encodes two distinct isoforms. (A) Alignment of the predicted NH<sub>2</sub>-terminal amino acid sequence of the p70 isoform of MGR and BOM. Amino acid identity is denoted by shared upper case letters and similarity by the (+) symbol. The first amino acids shared between p61 MGR and p70 MGR are given in bold. (B) Structure of the human and murine *mgr* genomic loci. Human genomic sequence was downloaded from the GenBank database (Accession Number AC010969) and aligned with cDNA sequences. Murine genomic clones were obtained from a 129 library and mapped by Southern analysis and PCR. Exons are denoted as E1-8 in human and E1-9 in murine. The two human MGR isoforms are denoted as p70 and p49 MGR and the two murine isoforms as p70 and p61 MGR. The scale of 1 kb is shown. (C) Identification of the murine p61 MGR promoter. Sequence was obtained from intron three from the MGR genomic locus and analyzed using the weight matrices of Bucher, *J. Mol. Biol.* 212: 563-578, 1990. The CAP site, TATA box and GC box are indicated. The cDNA start site is shown in arrows, the first ATG is given in bold and the splice site at the end of the first exon of p61 MGR is indicated.

Figure 2 is a photographic representation showing that p70 MGR binds to *Drosophila* gene regulatory sequences which bind *grh*. (A) p70 MGR binds to the *Drosophila* PCNA promoter. Nuclear extract from the JEG-3 cell line was studied in an EMSA with a PCNA promoter probe in the presence and absence of anti-MGR specific antisera. Antisera 611 was raised against peptides common to the p70 and p49 MGR proteins in the dimerization domain and antisera 67 was raised against unique peptides in the NH<sub>2</sub>-terminal domain of p70 MGR. The migration of the MGR complex is shown in arrows. (B) p70 MGR binds to the *Drosophila* *dopo decarboxylase* promoter. Experimental conditions were as described for (A).

Figure 3 are representations showing that p70 MGR binds to and transactivates the human En-1 promoter. (A) Identification of a *grh* consensus DNA binding site in the human En-1 promoter. The consensus sequence for *grh* DNA binding compiled from an alignment of



- 13 -

the *Drosophila Ultrabithorax*, *Dopa decarboxylase* and *fushi tarazu* promoters was compared with the sequence of the proximal human En-1 promoter and the *Drosophila engrailed* promoter. The closed bracket indicates the extend of the *grainyhead* binding site in the *engrailed* promoter as defined by DNaseI footprinting. (B) Human p70 MGR binds to the human En-1 promoter. Nuclear extract from the JEG-3 cell line was studied in an EMSA with a *Ddc* promoter probe in the presence of pre-immune sera (lane 1), anti-MGR specific antisera 67 (detailed in legend to Figure 2) (lane 2) or cold competitor DNA (lanes 3-5). A 50-fold excess of the *Ddc* probe was used in lane 3 and a 10- and 20-fold excess of a human En-1 promoter probe in lanes 4 and 5, respectively. The migration of the MGR/DNA complex is shown by arrows. (C) Human p70 MGR transactivates the En-1 promoter. COS cells were transiently transfected with the proximal En-1 promoter containing the MGR binding site linked to a minimal  $\gamma$ -globin promoter and a firefly luciferase reporter gene (solid columns), the minimal  $\gamma$ -globin promoter/luciferase reporter gene (open columns) and the TK promoter linked to the Renilla luciferase reporter gene (hatched columns) in the presence and absence of a p70 MGR expression vector (PCI-p70 MGR) as indicated. Transfection with the empty vector (pCI) served as the control. Luciferase levels were corrected for protein concentration and values were derived from two independent experiments performed in triplicate.

**Figure 4** is a photographic representation showing expression of GRHL-3 from E8.5 to E15.5. Sections of murine embryos were analysed by *in situ* hybridisation with a GRHL-3-specific  $^{33}\text{P}$ -labelled antisense riboprobe. (A) Transverse section of an E8 embryo showing two discrete areas of intense expression in non-neuronal ectoderm (arrowed) adjacent to the folding neural plate (bottom panel). The section counter-stained with hematoxylin is shown (top panel). da, dorsal aorta; hd, hind-gut diverticula; ne, neural epithelium; se, surface ectoderm. (B) Transverse section of E8 embryo probed with the control sense riboprobe (bottom panel) and the hematoxylin counter-stain (top panel). (C,D) Saggital sections of E12.5 (C) and E15.5 (D) embryos showing increasingly intense hybridisation to surface ectoderm (C and D). Hybridisation is also noted to other tissues lined by squamous epithelium including oral cavity, urogenital sinus and anal canal (D). ac, anal canal; dea, descending aorta; dv, ductus venosus; gt, genital tubercle; he, heart; li, liver; nc, nasal

- 14 -

cavity; np, nasal process; o, oral cavity; se, surface ectoderm; ta, tail; to, tongue; us, urogenital sinus. Signal from the descending aorta and ductus venosus is non-specific due to reflection from retained erythrocytes.

5 **Figure 5** is a representation showing the generation of a null allele of GRHL-3. (A) Gene-targeting strategy applied to the mouse GRHL-3 locus. The homologous recombination event deleted 2.2 kb of genomic DNA, including the region encoding the entire transcriptional activation domain of the protein. This was replaced with a promoter-less lacZ.polyA cassette fused to the second codon of exon 2 and a Neo<sup>R</sup> gene linked to a PGK

10 promoter and flanked by loxP sites. The thymidine kinase gene driven off the MC1-promoter completed the targeting vector. The location of the 5' and 3' probes used for Southern blot analysis of the targeted allele and the size of the expected hybridisation fragments prior to excision of the Neo<sup>R</sup> cassette are shown. The Neo<sup>R</sup> cassette was excised by crossing mice heterozygous for the targeted allele with a transgenic line expressing the

15 Cre recombinase gene driven off a CMV-promoter. LacZ.polyA, the lacZ gene linked to the rabbit  $\beta$ -globin polyadenylation signal; B, *Bam*HI; S, *Spe*I. (B) Southern blot analysis of two targeted ES cell clones (C7 and B12) and the parental ES cells (G7) with the 5' flanking probe demonstrating site-specific integration by homologous recombination. The size of DNA standards (in kb) is indicated. (C) Germ-line transmission of the targeted

20 allele from cell line C7. Southern blotting was performed with the 3' flanking probe on tail DNA isolated from weaned progeny of GRHL-3<sup>+/-</sup> intercrosses. The size of DNA standards (in kb) is indicated. (D) PCR genotyping of embryos. Two allele, three primer PCR was performed on genomic DNA from E18.5 embryos isolated from GRHL-3<sup>+/-</sup> intercrosses. The size of DNA standards (in bp) is indicated. target, PCR product diagnostic of targeted

25 GRHL-3 allele; wt, PCR product diagnostic of wild type GRHL-3 allele. (E) Northern blot analysis of GRHL-3 mRNA expression in wild type embryos and embryos heterozygous or homozygous for the targeted GRHL-3 allele (upper panel). RNA integrity was confirmed with a GAPDH probe (lower panel). The size of RNA standards is indicated, as is the migration of the GRHL-3 and GAPDH transcripts. (F) RT-PCR of E9.5 GRHL-3<sup>-/-</sup> and

30 GRHL-3<sup>+/-</sup> embryos was performed with primers specific for HPRT. Based on the HPRT quantitation, comparable amounts of cDNA from each embryo were PCR amplified for 30,

- 15 -

32 and 35 cycles (GRHL-3<sup>+/+</sup>) or 35, 38 and 40 cycles (GRHL-3<sup>-/-</sup>) with primers specific for GRHL-3. The 5' primer anneals to exon 8 and the 3' primer anneals to exon 13. Both primer pairs gave predicted size bands of 503 bp for GRHL-3 and 229 bp for HPRT. The identities of the amplified bands were confirmed by Southern blotting using gene-specific  
5 internal oligonucleotides.

**Figure 6** is a photographic representation of the phenotype of the GRHL-3-deficient mice. (A) E18.5 littermate embryos, wild type (+/+) and deficient (-/-). The range of NTDs in the GRHL-3<sup>-/-</sup> embryos are illustrated; exencephaly (arrow) and thoraco-lumbo-sacral spina bifida (arrowhead). Curled tails and growth retardation are also apparent in these embryos. A magnified view of the curly tail (ct) and the spina bifida from a caudal longitudinal view (clv) are inset. (B,C) Alizarin red/Alcian blue stained full-body skeletal preparations of E18.5 littermates illustrating the kyphosis (k) and tail flexion deformity (ct) in (B), and the abnormal vertebral pedicles in the thoraco-lumbo-sacral regions of the GRHL-3<sup>-/-</sup> embryo  
10 in (B and C). np, normal pedicles; sp, splayed pedicles. (D) transverse sections through +/- and -/- ES.5 to E14.5 embryos in the region of the caudal neural tube stained with hematoxylin and eosin. The open neural plate is arrowed.

**Figure 7** is a representation showing GRHL-3 and *ct* are the same gene. (A) Organization  
20 of the *ct* candidate region. Genetic map of the 13 Mb supercontig (Accession number NW\_000213) that shows the positions of relevant markers (D4Mit69 and D4Mit157) and previously excluded *ct* candidate genes (*Synd3*, *Fgr*, *Hspg2*, *Pax7*). The position of the GRHL-3 locus is also indicated. The size of the interval between the GRHL-3 locus and the D4Mit69 marker is shown. (B) Morphological appearance and genotype of embryos derived from *ct/ct* mice crossed with GRHL-3<sup>+/+</sup> mice. Embryo 1 is unremarkable; embryos 2 and 3 display curly tails (arrowheads); embryos 4 and 5 display curly tails and lumbo-sacral spina bifida (arrows). ct, curly tail; SB, spina bifida. Scale bar = 10mm. (C) Total RNA from E14.5 embryos from *curly tail* (*ct/ct*), wild type (+/+) and GRHL-3 heterozygotes (+/-) were analysed for GRHL-3 expression by Northern blotting with a  
30 cDNA probe derived from the unique coding portion of the mRNA described in Fig 1A (upper panel). RNA loading was monitored by probing with 28S (lower panel). Signal

- 16 -

intensity was quantified by Phosphorimager densitometry and the individual embryo GRHL-3 signals corrected for 28S loading. The corrected signal intensities relative to wild type embryo 7 are shown. Positions of GRHL-3, 28S and the RNA size standards are indicated. (D) Quantitative real-time RT-PCR was performed on total RNA from E14.5

5 *curly tail* (*ct/ct*), wild type (+/+) and GRHL-3 heterozygous (+/-) embryos. A standard curve was generated for HPRT and GRHL-3 and the relative quantity of both transcripts was calculated for individual embryos. Each reaction was performed in duplicate. The ratios of GRHL-3/HPRT in GRHL-3<sup>+/-</sup> and *ct/ct* embryos were normalised to the values obtained with GRHL-3<sup>+/+</sup> embryos.

- 17 -

**DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS**

The present invention is predicated in part on the identification of mammalian homologs of the *Drosophila* transcription factor known as Grainyhead (GRH). GRH is encoded by the gene, *grainyhead* (*grh*). In *Drosophila*, mutations in this gene are associated with embryonic lethal phenotypes, indicating the importance of the gene for normal development and function. The mammalian homologs are proposed to be involved in the regulation of developmental and/or non-developmental genes. Identification and isolation of the mammalian homologs of *grh* (M-*grh*) enable the development of a range of diagnostic and therapeutic agents useful in the detection and treatment of genetic disorders.

The present invention provides, therefore, a family of mammalian-derived transcription factors, highly related from *Drosophila* to mammals. These transcription factors are more highly conserved than CP2, LBP-1a and LBP-9. The present invention does not extend to CP2, LBP-1 and LBP-9. Reference to a mammal in this context includes a human, livestock animal (e.g. sheep, cow, horse, pig, donkey, goat), laboratory test animal (e.g. mouse, rat, rabbit, guinea pig), companion animal (e.g. dog, cat) or captive wild animal. Most preferably, the animal is a human or murine species. Sources of the isolated nucleic acid molecules include a range of tissues, such as mouse embryo, human fetal brain and placenta, and mouse and human kidney. In view of the highly conserved nature of this family of M-*grh* nucleotide sequences, however, corresponding homologs from other tissues and from other mammalian species are intended to be included within the scope of the present invention. The term "homolog" as used herein, therefore, extends to encompass transcription factors from mammalian species encoded by nucleotide sequences which have substantial similarity to *Drosophila* *grh* or a conserved region thereof. At the protein level, a homolog includes an amino acid sequence and/or tertiary structure having similarity to *Drosophila* GRH. In cases where the expression product of the M-*grh* is RNA, a homolog is defined by reference to the similar ribonucleotide sequence to that encoded by *Drosophila* *grh*.

- 18 -

M-grhd or M-GRH, i.e. a mammalian homolog of *Drosophila grh* or GRH is defined as such by having a nucleotide or amino acid sequence which has 60% or greater similarity after optimal alignment to *Drosophila grh* or GRH.

- 5 Accordingly, one aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a mammalian homolog of *Drosophila grh*.

Reference to a mammalian homolog of *Drosophila* GRH (i.e. a M-GRH) preferably  
10 includes the mammalian homolog of grainyhead (MGR), brother of MGR (BOM) and sister of MGR (SOM). These transcription factors are encoded by *mgr*, *bom* and *som*, respectively. Reference to "MGR", "BOM" and "SOM" or *mgr*, *bom* and *som* includes all mutants, derivatives, homologs and analogs thereof. The present invention further extends, however, to all novel mammalian homologs of *Drosophila grh* but does not encompass  
15 CP2, LBP-1a or LBP-9. The nucleotide sequences for *Drosophila grh* are set forth in SEQ ID NO:17, SEQ ID NO:34, SEQ ID NO:36 and SEQ ID NO:38, respectively. Consequently, a mammalian homolog is defined herein as comprising a nucleotide sequence having at least about 60% sequence similarity to SEQ ID NO:17 or SEQ ID NO:34 or SEQ ID NO:36 or SEQ ID NO:38 after optimal alignment and/or being capable  
20 of hybridizing to SEQ ID NO:17 or SEQ ID NO:34 or SEQ ID NO:36 or SEQ ID NO:38 or its complementary form under low stringency conditions.

Accordingly, another aspect of the present invention provides an isolated nucleic acid molecule encoding a mammalian transcription factor or a functional part thereof  
25 comprising a sequence of nucleotides having at least 60% similarity to SEQ ID NO:17 or SEQ ID NO:34 or SEQ ID NO:36 or SEQ ID NO:38 after optimal alignment and/or being capable of hybridizing to SEQ ID NO:17 or its complementary form under low stringency conditions.

30 In a preferred embodiment, the isolated nucleic acid molecule encodes a proteinaceous form of a transcription factor. Examples of such mammalian protein transcription factors

- 19 -

include human MGR p49 (SEQ ID NO:2), human MGR p70 (SEQ ID NO:4), human BOM (SEQ ID NO:6), human SOM (SEQ ID NO:7), murine MGR p61 (SEQ ID NO:10), murine MGR p70 (SEQ ID NO:12), murine BOM (SEQ ID NO:14) and murine SOM (SEQ ID NO:16).

5

Accordingly, another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding a polypeptide having transcription factor activity and comprising an amino acid sequence substantially as set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ  
10 ID NO:12, SEQ ID NO:14 or SEQ ID NO:16 or an amino acid sequence having at least about 60% similarity to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 or SEQ ID NO:16 after optimal alignment wherein said polypeptide is a mammalian homolog of *Drosophila* GRH.

15 Such a polypeptide is referred to herein as a M-GRH.

Preferred percentage amino acid similarity levels include at least about 61% or at least about 62% or at least about 63% or at least about 64% or at least about 65% or at least about 66% or at least about 67% or at least about 68% or at least about 69% or at least  
20 about 70% or at least about 71% or at least about 72% or at least about 73% or at least about 74% or at least about 75% or at least about 76% or at least about 77% or at least about 78% or at least about 79% or at least about 80% or at least about 81% or at least about 82% or at least about 83% or at least about 84% or at least about 85% or at least about 86% or at least about 87% or at least about 88% or at least about 89% or at least  
25 about 90% or at least about 91% or at least about 92% or at least about 93% or at least about 94% or at least about 95% or at least about 96% or at least about 97% or at least about 98% or at least about 99% similarity.

This aspect of the present invention includes derivatives of M-GRH molecules. Such  
30 derivatives include non-active fragments which encompass *inter alia* the binding domain as well as active isoforms.

- 20 -

A "derivative" of a polypeptide of the present invention also encompasses a portion or a part of a full-length parent polypeptide, which retains the transcription factor activity of the parent polypeptide. Such "biologically-active fragments" include deletion mutants and small peptides, for example, of at least 10, preferably at least 20 and more preferably at least 30 contiguous amino acids, which exhibit the requisite activity. Peptides of this type may be obtained through the application of standard recombinant nucleic acid techniques or synthesized using conventional liquid or solid phase synthesis techniques. For example, reference may be made to solution synthesis or solid phase synthesis as described, for example, in Chapter 9 entitled "*Peptide Synthesis*" by Atherton and Shephard which is included in a publication entitled "*Synthetic Vaccines*" edited by Nicholson and published by Blackwell Scientific Publications. Alternatively, peptides can be produced by digestion of an amino acid sequence of the invention with proteinases such as endoLys-C, endoArg-C, endoGlu-C and staphylococcus V8-protease. The digested fragments can be purified by, for example, high performance liquid chromatographic (HPLC) techniques. Any such fragment, irrespective of its means of generation, is to be understood as being encompassed by the term "derivative" as used herein.

In another embodiment, the present invention provides an isolated nucleic acid molecule encoding a mammalian transcription factor homolog of *Drosophila grh* (i.e. a M-GRH) and comprising a nucleotide sequence selected from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 and SEQ ID NO:15 or a nucleotide sequence having at least about 60% similarity to any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 after optimal alignment or a nucleotide sequence capable of hybridizing to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or a complementary form thereof under low stringency conditions.

Preferably, percentage nucleotide similarity levels include at least about 61% or at least about 62% or at least about 63% or at least about 64% or at least about 65% or at



- 21 -

least about 66% or at least about 67% or at least about 68% or at least about 69% or at least about 70% or at least about 71% or at least about 72% or at least about 73% or at least about 74% or at least about 75% or at least about 76% or at least about 77% or at least about 78% or at least about 79% or at least about 80% or at least about 81% or at least about 82% or at least about 83% or at least about 84% or at least about 85% or at least about 86% or at least about 87% or at least about 88% or at least about 89% or at least about 90% or at least about 91% or at least about 92% or at least about 93% or at least about 94% or at least about 95% or at least about 96% or at least about 97% or at least about 98% or at least about 99% similarity.

10

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. In a particularly preferred embodiment, nucleotide and sequence comparisons are made at the level of identity rather than similarity.

20

Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include "reference sequence", "comparison window", "sequence similarity", "sequence identity", "percentage of sequence similarity", "percentage of sequence identity", "substantially similar" and "substantial identity". A "reference sequence" is at least 12 but frequently 15 to 18 and often at least 25 or above, such as 30 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (i.e. only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of

30

- 22 -

sequence similarity. A "comparison window" refers to a conceptual segment of typically 12 contiguous residues that is compared to a reference sequence. The comparison window may comprise additions or deletions (i.e. gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (i.e. resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.* (*Nucl. Acids. Res.* 25: 3389, 1997). A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.* (In: Current Protocols in Molecular Biology, John Wiley & Sons Inc. 1994-1998).

The terms "sequence similarity" and "sequence identity" as used herein refers to the extent that sequences are identical or functionally or structurally similar on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity", for example, is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g. A, T, C, G, I) or the identical amino acid residue (e.g. Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For the purposes of the present invention, "sequence identity" will be understood to mean the "match percentage" calculated by the DNASIS computer program (Version 2.5 for windows; available from Hitachi Software engineering Co., Ltd., South San Francisco, California, USA) using standard defaults as used in the reference manual accompanying the software. Similar comments apply in relation to sequence similarity.

- 23 -

The present invention provides, therefore, an isolated nucleic acid molecule comprising a sequence of nucleotides selected from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 and SEQ ID NO:15 or a complementary form thereof. Such nucleic acid molecules encode mammalian homologs of *Drosophila grh*. These mammalian homologs are proposed herein to be transcription factors.

The present invention extends to variants of the nucleic acid molecules. A variant is a molecule having less than 100% sequence identity to a M-*grh*. Generally, a variant will still hybridize to a M-*grh* sequence under low stringency conditions.

The term "variant" refers, therefore, to nucleotide sequences displaying substantial sequence identity with a reference nucleotide sequences or polynucleotides that hybridize with a reference sequence under stringency conditions that are defined hereinafter. The terms "nucleotide sequence", "polynucleotide" and "nucleic acid molecule" may be used herein interchangeably and encompass polynucleotides in which one or more nucleotides have been added or deleted, or replaced with different nucleotides. In this regard, it is well understood in the art that certain alterations inclusive of mutations, additions, deletions and substitutions can be made to a reference nucleotide sequence whereby the altered polynucleotide retains the biological function or activity of the reference polynucleotide. The term "variant" also includes naturally-occurring allelic variants.

Reference herein to a low stringency includes and encompasses from at least about 0 to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridization, and at least about 1 M to at least about 2 M salt for washing conditions. Generally, low stringency is at from about 25-30°C to about 42°C. The temperature may be altered and higher temperatures used to replace formamide and/or to give alternative stringency conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridization, and at least about 0.5 M to at least about 0.9 M salt for washing

- 24 -

conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01 M to at least about 0.15 M salt for hybridization, and at least about 0.01 M to at least about 0.15 M salt for washing conditions. In general, washing is carried out  $T_m = 69.3 + 0.41 (G+C)\%$  (Marmur and Doty, *J. Mol. Biol.* 5: 109, 1962). However, the  $T_m$  of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatch base pairs (Bonner and Laskey, *Eur. J. Biochem.* 46: 83, 1974). Formamide is optional in these hybridization conditions. Accordingly, particularly preferred levels of stringency are defined as follows: low stringency is 6 x SSC buffer, 0.1% w/v SDS at 25°-42°C; a moderate stringency is 2 x SSC buffer, 0.1% w/v SDS at a temperature in the range 20°C to 65°C; high stringency is 0.1 x SSC buffer, 0.1% w/v SDS at a temperature of at least 65°C.

The present invention extends to recombinant forms of the M-*grh* molecules as well as derivatives and homologs thereof.

15

Accordingly, another aspect of the present invention provides an isolated polypeptide having transcription factor activity, said polypeptide comprising a sequence of amino acids encoded by a nucleotide sequence selected from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or a nucleotide sequence having at least about 60% similarity to any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or a nucleotide sequence capable of hybridizing to any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or a complementary form thereof under low stringency conditions.

25

In a preferred embodiment, the present invention provides a recombinant M-*grh* comprising an amino acid sequence selected from SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 or SEQ ID NO:16 or an amino acid sequence having at least about 60% similarity to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 or

30

SEQ ID NO:16.

This aspect of the present invention extends to derivatives, homologs and analogs of M-GRH molecules.

5

A "derivative" includes a mutant, fragment, part, portion or hybrid molecule. A derivative generally but not exclusively carries a single or multiple amino acid substitution, addition and/or deletion.

10 A "homolog" includes an analogous polypeptide having at least about 60% similar amino acid sequence from another animal species or from a different locus within the same species.

An "analog" is generally a chemical analog. Chemical analogs of the subject polypeptide  
15 contemplated herein include, but are not limited to, modification to side chains, incorporation of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogs.

20 Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with  $\text{NaBH}_4$ ; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS);  
25 acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with  $\text{NaBH}_4$ .

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal  
30 and glyoxal.

- 26 -

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitization, for example, to a corresponding amide.

5 Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline  
10 pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with  
15 tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

20

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl  
25 alanine and/or D-isomers of amino acids. A list of unnatural amino acid, contemplated herein is shown in Table 3.

TABLE 3

	Non-conventional amino acid	Code	Non-conventional amino acid	Code
5	$\alpha$ -aminobutyric acid	Abu	L-N-methylalanine	Nmala
	$\alpha$ -amino- $\alpha$ -methylbutyrate	Mgab	L-N-methylarginine	Nmarg
	aminocyclopropane-	Cpro	L-N-methylasparagine	Nmasn
10	carboxylate		L-N-methylaspartic acid	Nmasp
	aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
	aminonorbornyl-	Norb	L-N-methylglutamine	Nmglu
	carboxylate		L-N-methylglutamic acid	Nmglu
	cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
15	cyclopentylalanine	Cpen	L-N-methylisoleucine	Nmile
	D-alanine	Dal	L-N-methylleucine	Nmleu
	D-arginine	Darg	L-N-methyllysine	Nmlys
	D-aspartic acid	Das	L-N-methylmethionine	Nmmet
	D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
20	D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
	D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
	D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
	D-isoleucine	Dile	L-N-methylproline	Nmpro
	D-leucine	Dleu	L-N-methylserine	Nmser
25	D-lysine	Dlys	L-N-methylthreonine	Nmthr
	D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
	D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
	D-phenylalanine	Dphe	L-N-methylvaline	Nmval
	D-proline	Dpro	L-N-methylethylglycine	Nmetg
30	D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
	D-threonine	Dthr	L-norleucine	Nle

- 28 -

	D-tryptophan	Dtrp	L-norvaline	Nva
	D-tyrosine	Dtyr	$\alpha$ -methyl-aminoisobutyrate	Maib
	D-valine	Dval	$\alpha$ -methyl- $\gamma$ -aminobutyrate	Mgab
	D- $\alpha$ -methylalanine	Dmala	$\alpha$ -methylcyclohexylalanine	Mchexa
5	D- $\alpha$ -methylarginine	Dmarg	$\alpha$ -methylcyclopentylalanine	Mcpen
	D- $\alpha$ -methylasparagine	Dmasn	$\alpha$ -methyl- $\alpha$ -naphthylalanine	Manap
	D- $\alpha$ -methylaspartate	Dmasp	$\alpha$ -methylpenicillamine	Mpen
	D- $\alpha$ -methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D- $\alpha$ -methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
10	D- $\alpha$ -methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
	D- $\alpha$ -methylisoleucine	Dmile	N-amino- $\alpha$ -methylbutyrate	Nmaabu
	D- $\alpha$ -methylleucine	Dmleu	$\alpha$ -naphthylalanine	Anap
	D- $\alpha$ -methyllysine	Dmlys	N-benzylglycine	Nphe
	D- $\alpha$ -methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
15	D- $\alpha$ -methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
	D- $\alpha$ -methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D- $\alpha$ -methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D- $\alpha$ -methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D- $\alpha$ -methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
20	D- $\alpha$ -methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
	D- $\alpha$ -methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	D- $\alpha$ -methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
25	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nblhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
30	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl)glycine	Nser



- 29 -

	D-N-methylisoleucine	Dnmile	N-(imidazolyethyl)glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp
	D-N-methyllysine	Dnmlys	N-methyl- $\gamma$ -aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
5	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
10	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyl- $\alpha$ -naphthylalanine	Nmanap
	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	$\gamma$ -aminobutyric acid	Gabu	N-( <i>p</i> -hydroxyphenyl)glycine	Nhtyr
	L- <i>t</i> -butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
15	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L- $\alpha$ -methylalanine	Mala
	L- $\alpha$ -methylarginine	Marg	L- $\alpha$ -methylasparagine	Masn
	L- $\alpha$ -methylaspartate	Masp	L- $\alpha$ -methyl- <i>t</i> -butylglycine	Mtbug
	L- $\alpha$ -methylcysteine	Mcys	L-methylethylglycine	Metg
20	L- $\alpha$ -methylglutamine	Mgln	L- $\alpha$ -methylglutamate	Mglu
	L- $\alpha$ -methylhistidine	Mhis	L- $\alpha$ -methylhomophenylalanine	Mhphe
	L- $\alpha$ -methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L- $\alpha$ -methylleucine	Mleu	L- $\alpha$ -methyllysine	Mlys
	L- $\alpha$ -methylmethionine	Mmet	L- $\alpha$ -methylnorleucine	Mnle
25	L- $\alpha$ -methylnorvaline	Mnva	L- $\alpha$ -methylornithine	Morn
	L- $\alpha$ -methylphenylalanine	Mphe	L- $\alpha$ -methylproline	Mpro
	L- $\alpha$ -methylserine	Mser	L- $\alpha$ -methylthreonine	Mthr
	L- $\alpha$ -methyltryptophan	Mtrp	L- $\alpha$ -methyltyrosine	Mtyr
	L- $\alpha$ -methylvaline	Mval	L-N-methylhomophenylalanine	Nmhph
30	N-(N-(2,2-diphenylethyl)	Nnbhm	N-(N-(3,3-diphenylpropyl)	Nnbhe

- 30 -

carbamylmethyl)glycine

carbamylmethyl)glycine

1-carboxy-1-(2,2-diphenyl- Nmbc

ethylamino)cyclopropane

5

Crosslinkers can be used, for example, to stabilize 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having  $(CH_2)_n$  spacer groups with  $n=1$  to  $n=6$ , glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of  $C_\alpha$  and N  $\alpha$ -methylamino acids, introduction of double bonds between  $C_\alpha$  and  $C_\beta$  atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

The present invention further contemplates chemical analogs of the subject polypeptide capable of acting as antagonists or agonists of M-GRH or which can act as functional analogs of M-GRH. Chemical analogs may not necessarily be derived from the instant M-GRH molecules but may share certain conformational similarities. Alternatively, chemical analogs may be specifically designed to mimic certain physiochemical properties of the subject M-GRH molecules. Chemical analogs may be chemically synthesized or may be detected following, for example, natural product screening. The latter refers to molecules identified from various environmental sources such as river beds, coral, plants, microorganisms and insects.

These types of modifications may be important to stabilize the subject M-GRH molecules if administered to an individual or for use as a diagnostic reagent.

- 31 -

Other derivatives contemplated by the present invention include a range of glycosylation variants from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

5

The designing of mimetics to a pharmaceutically active compound is a known approach to the development of pharmaceuticals based on a "lead" compound. This might be desirable where the active compound is difficult or expensive to synthesize or where it is unsuitable for a particular method of administration, e.g. peptides are unsuitable active agents for oral compositions as they tend to be quickly degraded by proteases in the alimentary canal. Mimetic design, synthesis and testing is generally used to avoid randomly screening large numbers of molecules for a target property.

10

There are several steps commonly taken in the design of a mimetic from a compound having a given target property. First, the particular parts of the compound that are critical and/or important in determining the target property are determined. In the case of a peptide, this can be done by systematically varying the amino acid residues in the peptide, e.g. by substituting each residue in turn. Alanine scans of peptides are commonly used to refine such peptide motifs. These parts or residues constituting the active region of the compound are known as its "pharmacophore".

15

20

Once the pharmacophore has been found, its structure is modeled according to its physical properties, e.g. stereochemistry, bonding, size and/or charge, using data from a range of sources, e.g. spectroscopic techniques, x-ray diffraction data and NMR. Computational analysis, similarity mapping (which models the charge and/or volume of a pharmacophore, rather than the bonding between atoms) and other techniques can be used in this modeling process.

25

In a variant of this approach, the three-dimensional structure of the ligand and its binding partner are modeled. This can be especially useful where the ligand and/or binding partner change conformation on binding, allowing the model to take account of this in the design

30

- 32 -

of the mimetic. Modeling can be used to generate inhibitors which interact with the linear sequence or a three-dimensional configuration.

5 A template molecule is then selected onto which chemical groups which mimic the pharmacophore can be grafted. The template molecule and the chemical groups grafted onto it can conveniently be selected so that the mimetic is easy to synthesize, is likely to be pharmacologically acceptable, and does not degrade *in vivo*, while retaining the biological activity of the lead compound. Alternatively, where the mimetic is peptide-based, further stability can be achieved by cyclizing the peptide, increasing its rigidity. The mimetic or  
10 mimetics found by this approach can then be screened to see whether they have the target property, or to what extent they exhibit it. Further optimization or modification can then be carried out to arrive at one or more final mimetics for *in vivo* or clinical testing.

The goal of rational drug design is to produce structural analogs of biologically active  
15 polypeptides of interest or of small molecules with which they interact (e.g. agonists, antagonists, inhibitors or enhancers) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g. enhance or interfere with the function of a polypeptide *in vivo*. See, e.g. Hodgson (*BioTechnology* 9: 19-21, 1991). In one approach, one first determines the three-dimensional structure of a protein of interest  
20 by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Useful information regarding the structure of a polypeptide may also be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson *et al.*, *Science* 249: 527-533, 1990). In addition, target molecules may be analyzed by an alanine scan (Wells,  
25 *Methods Enzymol.* 202: 2699-2705, 1991). In this technique, an amino acid residue is replaced by Ala and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

- 33 -

It is also possible to isolate a target-specific antibody, selected by a functional assay and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional,  
5 pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

- 10 Two-hybrid screening is also useful in identifying other members of a biochemical or genetic pathway associated with a target. Two-hybrid screening conveniently uses *Saccharomyces cerevisiae* and *Saccharomyces pombe*. Target interactions and screens for inhibitors can be carried out using the yeast two-hybrid system, which takes advantage of transcriptional factors that are composed of two physically separable, functional domains.
- 15 The most commonly used is the yeast GAL4 transcriptional activator consisting of a DNA binding domain and a transcriptional activation domain. Two different cloning vectors are used to generate separate fusions of the GAL4 domains to genes encoding potential binding proteins. The fusion proteins are co-expressed, targeted to the nucleus and if interactions occur, activation of a reporter gene (e.g. *lacZ*) produces a detectable  
20 phenotype. In the present case, for example, *S. cerevisiae* is co-transformed with a library or vector expressing a cDNA GAL4 activation domain fusion and a vector expressing a holocyclotxin-GAL4 binding domain fusion. If *lacZ* is used as the reporter gene, co-expression of the fusion proteins will produce a blue color. Small molecules or other candidate compounds which interact with a target will result in loss of colour of the cells.
- 25 Reference may be made to the yeast two-hybrid systems as disclosed by Munder *et al.* (*Appl. Microbiol. Biotechnol.* 52: 311-320, 1999) and Young *et al.* (*Nat. Biotechnol.* 16: 946-950, 1998). Molecules thus identified by this system are then re-tested in animal cells.

The present invention further contemplates methods of screening for drugs comprising, for  
30 example, contacting a candidate drug with a transcription factor. These molecules are referred to herein as "targets", "a target" or "target molecule". The screening procedure

- 34 -

includes assaying for the presence of a complex between the drug and the target. One form of assay involves competitive binding assays. In such competitive binding assays, the target is typically labeled. Free target is separated from any putative complex and the amount of free (i.e. uncomplexed) label is a measure of the binding of the agent being tested to target molecule. One may also measure the amount of bound, rather than free, target. It is also possible to label the compound rather than the target and to measure the amount of compound binding to target in the presence and in the absence of the drug being tested.

- 10 Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to a target and is described in detail in Geysen (International Patent Publication No. WO 84/03564). Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with a target and washed. Bound target molecule is then detected by methods well known in the art. This method may be adapted for screening for non-peptide, chemical entities. This aspect, therefore, extends to combinatorial approaches to screening for target antagonists or agonists.
- 20 Purified target can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the target may also be used to immobilize the target on the solid phase.

25 The present invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the target compete with a test compound for binding to the target or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the target.

- 35 -

The present invention also provides a method for identifying a M-GRH, said method comprising screening a nucleotide database and identifying a nucleotide sequence having at least 60% similarity to SEQ ID NO:17 or SEQ ID NO:34 or SEQ ID NO:36 or SEQ ID NO:38 after optimal alignment.

5

Reference to a "nucleotide database" includes screening an existing genomic or cDNA or mRNA database or screening for a target nucleic acid molecule in a mammalian cell such as using oligonucleotide probes or primers, sequencing the target molecule and comparing the sequence to SEQ ID NO:17 or SEQ ID NO:34 or SEQ ID NO:36 or SEQ ID NO:38.

10

In an alternative method, a database of mammalian protein sequences is screened for an amino acid sequence having at least 60% similarity to the amino acid sequence encoded by SEQ ID NO:17 or SEQ ID NO:34 or SEQ ID NO:36 or SEQ ID NO:38. Again, a "database" includes a *de novo* protein sequence isolated and identified on a transcription factor isolated from a mammalian cell.

15

In yet another alternative, a M-*grh* or its protein product is deemed one which has at least about 60% similarity at the nucleotide level to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or at the amino acid level to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 or SEQ ID NO:15.

20

Still yet another aspect of the present invention provides a means of identifying a nucleotide sequence likely to encode an M-GRH transcription factor, said method comprising interrogating a mammalian genome database conceptually translated into different reading frames with an amino acid sequence defining *Drosophila GRH* or any one of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 and SEQ ID NO:16 and identifying a nucleotide sequence corresponding to an amino acid sequence having at least about 60% similarity to *Drosophila GRH* or to any one of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 and SEQ ID NO:16.

30

- 36 -

Preferably, the genome is conceptually translated into from about 3 to about 6 reading frames and more preferably six reading frames.

5 It is proposed in accordance with the present invention that the M-GRH transcription factors are involved in the modulation of expression of a number of genes including developmentally regulated genes. Accordingly, aberrations in the M-GRH or M-*grh* molecules are proposed to cause over or under expression of particular genes leading to a potentially unwanted phenotype. The phenotype may manifest itself pre- or post-natally. A  
10 pre-natal manifestation includes at the embryo or fetus stage. Conditions contemplated include developmentally-determined disease conditions such as poor brain development, poor muscle or bone development, aberrations in facial or cranial structures, malformed spinal structures, predispositions to a range of cancers including melanomas and immunological disorders.

15 Accordingly, another aspect of the present invention contemplates a method for detecting an aberrant phenotype or a propensity for an aberrant phenotype to develop, said method comprising screening for a variation in a nucleotide sequence encoding a mammalian MGR, BOM and/or SOM or their homologs.

20 Reference herein to "MGR", "BOM" and "SOM" includes murine and human forms of these molecules such as human MGR p49 (SEQ ID NO:2), human MGR p70 (SEQ ID NO:4), human BOM (SEQ ID NO:6), human SOM (SEQ ID NO:8), murine MGR p61 (SEQ ID NO:10), murine MGR p70 (SEQ ID NO:12), murine BOM (SEQ ID NO:14) and  
25 murine SOM (SEQ ID NO:16).

A homolog of MGR, BOM and SOM is as herein defined including a molecule having at least about 60% amino acid sequence similarity to MGR, BOM or SOM or at least about 60% nucleic acid similarity to *mgr*, *bom* or *som* or a nucleic acid molecule capable of  
30 hybridizing to the coding strands of *mgr*, *bom* or *som* or complementary forms thereof under low stringency conditions.



- 37 -

Aberrations may also be detectable at the amino acid level when the mammalian homologs of *Drosophila grh* encode protein transcription factors.

5 Accordingly, another aspect of the present invention contemplates a method for detecting an aberrant phenotype or a propensity for an aberrant phenotype to develop, said method comprising screening for a variation in an amino acid sequence encoding MGR, BOM and/or SOM or their homologs.

10 As above, reference to MGR, BOM and SOM include amino acid sequences defining human MGR p49 (SEQ ID NO:2), human MGR p70 (SEQ ID NO:4), human BOM SEQ ID NO:6), human SOM (SEQ ID NO:8), murine MGR p61 (SEQ ID NO:10), murine MGR p70 (SEQ ID NO:12), murine BOM (SEQ ID NO:14) and murine SOM (SEQ ID NO:16).

15

As stated above, the mammalian transcription factors and their genetic sequences have a range of diagnostic and therapeutic utilities. The detection of an aberrant transcription factor or a nucleotide sequence encoding an aberrant transcription factor is indicative of a disease condition including a degenerative or developmental disease condition.

20

Any number of methods may be employed to detect aberrant transcription factors or their genetic sequences. Immunological testing is one particular method. Accordingly, the present invention extends to antibodies and other immunological agents directed to or preferably specific for the mammalian transcription factors or a fragment thereof. The  
25 antibodies may be monoclonal or polyclonal or may comprise Fab fragments or synthetic forms.

Specific antibodies can be used to screen for the subject mammalian transcription factors and/or their fragments. Techniques for the assays contemplated herein are known in the art  
30 and include, for example, sandwich assays and ELISA.

- 38 -

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies referred to above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of the mammalian transcription factors.

Both polyclonal and monoclonal antibodies are obtainable by immunization with mammalian transcription factors or antigenic fragments thereof and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal with an effective amount of subject polypeptide, or antigenic parts thereof, collecting serum from the animal and isolating specific sera by any of the known immunoabsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates, therefore, a method for detecting a mammalian transcription factor or fragment thereof in a biological sample from a subject, said method comprising contacting said biological sample with an antibody specific for said mammalian transcription factor or fragment thereof or its derivatives or homologs for a time and under conditions sufficient for an antibody-polypeptide complex to form, and then detecting said complex.

30

- 39. -

A biological sample includes a cell extract.

Reference to a "mammalian transcription factor" is considered to be a reference to a homolog of *Drosophila grh*, i.e. M-GRH.

5

The presence of the instant mammalian transcription factors or their fragments may be detected in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to U.S. Patent Nos. 4,016,043, 4,424,279 and 4,018,653.

10

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabeled antibody is immobilized on a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labeled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labeled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labeled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain a subject transcription factor including by tissue biopsy, blood, synovial fluid and/or lymph. The sample is, therefore, generally a biological sample comprising biological fluid. The transcription factor is likely to be in blood or other fluid in the case where cell apoptosis is occurring.

30

- 40 -

In the typical forward sandwich assay, a first antibody having specificity for the instant polypeptide or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or where more convenient, overnight) and under suitable conditions (e.g. for about 20°C to about 40°C) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labeled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labeled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules. In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however,

- 41 -

a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the  
5 corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labeled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution  
10 containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as  
15 red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labeled antibody  
20 adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. The fluorescent labeled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of  
25 the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

- 42 -

The present invention also contemplates genetic assays such as involving PCR analysis to detect RNA expression products of a genetic sequence encoding a mammalian transcription factor. The genetic assays may also be able to detect nucleotide polymorphisms or other substitutions, additions and/or deletions in the nucleotide sequence of a mammalian transcription factor. Changes in levels of mammalian transcription factor expression such as following mutations in the promoter or regulatory regions or loss of mammalian transcription factor activity following mutations in mammalian transcription factor nucleotides is proposed to be indicative of a disease condition or a propensity for a disease condition to develop. For example, a cell biopsy could be obtained and DNA or RNA extracted. Alternative methods which may be used alone or in conjunction with other methods include direct nucleotide sequencing or mutation scanning such as single stranded conformation polymorphisms analysis (SSCP) as well as specific oligonucleotide hybridization, denaturing high performance liquid chromatography, first nucleotide change (FNC) amongst others.

The present invention extends to polymorphisms which in the *M-grh* genes leads to healthy or abnormal phenotypes.

The present invention further contemplates kits to facilitate the rapid detection of mammalian transcription factors or their fragments in a subject's biological fluid.

Again, a biological fluid includes a cell extract such as a DNA/RNA extract.

Still yet another aspect of the present invention contemplates genomic sequences including gene sequences encoding a mammalian transcription factor as well as regulatory regions such as promoters, terminators and transcription/translation enhancer regions associated with the gene encoding a mammalian transcription factor.

- 43 -

The term "gene" is used in its broadest sense and includes cDNA corresponding to the exons of a gene. Accordingly, reference herein to a "gene" is to be taken to include:-

- (i) a classical genomic gene consisting of transcriptional and/or translational regulatory sequences and/or a coding region and/or non-translated sequences (i.e. introns, 5'- and 3'- untranslated sequences); or
- (ii) mRNA or cDNA corresponding to the coding regions (i.e. exons) and 5'- and 3'- untranslated sequences of the gene.

The term "gene" is also used to describe synthetic or fusion molecules encoding all or part of an expression product. In particular embodiments, the term "nucleic acid molecule" and "gene" may be used interchangeably.

In a particularly useful embodiment, the present invention provides a promoter for the mammalian transcription factor gene. The identification of the promoter permits developmentally-regulated expression of particular genetic sequences. The latter would include a range of therapeutic molecules such as cytokines, growth factors, antibiotics or other molecules to assist in the treatment of particular disease conditions.

Accordingly, another aspect of the present invention provides a *M-grh*-specific promoter or functional derivative or homolog thereof, said promoter *in situ* operably linked to a nucleotide sequence comprising any one of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or their complementary forms or a nucleotide sequence having at least about 60% similarity to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or their complementary forms or a nucleotide sequence capable of hybridizing to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or their complementary forms under low stringency conditions.

- 44 -

The promoter is conveniently resident in a vector which comprises unique restriction sites to facilitate the introduction of genetic sequences operably linked to the promoter.

All such constructs are useful in order to produce recombinant M-GRH molecules and/or  
5 in gene therapy protocols.

The present invention further contemplates a genetically modified animal.

More particularly, the present invention provides an animal model useful for screening for  
10 agents capable of ameliorating the effects of an aberrant M-GRH or M-*grh* gene. In one embodiment, the animal model produces low amounts of M-*grh*. Such an animal would have a predisposition for a range of diseases including developmentally regulated diseases. The animal model is useful for screening for agents which ameliorate such conditions.

15 Accordingly, another aspect of the present invention provides a genetically modified animal wherein said animal produces low amounts of M-*grh* relative to a non-genetically modified animal of the same species. Reference to "low amounts" includes zero amounts or up to about 10% lower than normalized amounts.

20 Preferably, the genetically modified animal is a mouse, rat, guinea pig, rabbit, pig, sheep or goat. More preferably, the genetically modified animal is a mouse or rat. Most preferably, the genetically modified animal is a mouse.

Accordingly, a preferred aspect of the present invention provides a genetically modified  
25 mouse wherein said mouse produces low amounts of M-*grh* relative to a non-genetically modified mouse of the same strain.

The animal model contemplated by the present invention comprises, therefore, an animal which is substantially incapable of producing a M-*grh*. Generally, but not exclusively, such  
30 an animal is referred to as a homozygous or heterozygous M-*grh*-knockout animal.



- 45 -

The animal models of the present invention may be in the form of the animals or may be, for example, in the form of embryos for transplantation. The embryos are preferably maintained in a frozen state and may optionally be sold with instructions for use.

- 5 The genetically modified animals may also produce larger amounts of M-GRH. For example, over expression of normal *M-grh* or mutant *M-grh* may produce dominant negative effects and may become useful disease models.

10 Accordingly, another aspect of the present invention is directed to a genetically modified animal over-expressing genetic sequences encoding *M-grh*.

A genetically modified animal includes a transgenic animal, or a "knock-out" or "knock-in" animal.

- 15 Yet another aspect of the present invention provides a targeting vector useful for inactivating a gene encoding M-GRH, said targeting vector comprising two segments of genetic material encoding said M-GRH flanking a positive selectable marker wherein when said targeting vector is transfected into embryonic stem (ES) cells and the marker selected, an ES cell is generated in which the gene encoding said M-GDH is inactivated by  
20 homologous recombination.

Preferably, the ES cells are from mice, rats, guinea pigs, pigs, sheep or goats. Most preferably, the ES cells are from mice.

- 25 Still yet another aspect of the present invention is directed to the use of a targeting vector as defined above in the manufacture of a genetically modified animal substantially incapable of producing M-GRH.

30 Even still another aspect of the present invention is directed to the use of a targeting vector as defined above in the manufacture of a genetically modified mouse substantially incapable of producing M-GRH.

- 46 -

Preferably, the vector is DNA. A selectable marker in the targeting vector allows for selection of targeted cells that have stably incorporated the targeting DNA. This is especially useful when employing relatively low efficiency transformation techniques such as electroporation, calcium phosphate precipitation and liposome fusion where typically fewer than 1 in 1000 cells will have stably incorporated the exogenous DNA. Using high efficiency methods, such as microinjection into nuclei, typically from 5-25% of the cells will have incorporated the targeting DNA; and it is, therefore, feasible to screen the targeted cells directly without the necessity of first selecting for stable integration of a selectable marker. Either isogenic or non-isogenic DNA may be employed.

Examples of selectable markers include genes conferring resistance to compounds such as antibiotics, genes conferring the ability to grow on selected substrates, genes encoding proteins that produce detectable signals such as luminescence. A wide variety of such markers are known and available, including, for example, antibiotic resistance genes such as the neomycin resistance gene (*neo*) and the hygromycin resistance gene (*hyg*). Selectable markers also include genes conferring the ability to grow on certain media substrates such as the *tk* gene (thymidine kinase) or the *hprt* gene (hypoxanthine phosphoribosyltransferase) which confer the ability to grow on HAT medium (hypoxanthine, aminopterin and thymidine); and the bacterial *gpt* gene (guanine/xanthine phosphoribosyltransferase) which allows growth on MAX medium (mycophenolic acid, adenine and xanthine). Other selectable markers for use in mammalian cells and plasmids carrying a variety of selectable markers are described in Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbour, New York, USA, 1990.

25

The preferred location of the marker gene in the targeting construct will depend on the aim of the gene targeting. For example, if the aim is to disrupt target gene expression, then the selectable marker can be cloned into targeting DNA corresponding to coding sequence in the target DNA. Alternatively, if the aim is to express an altered product from the target gene, such as a protein with an amino acid substitution, then the coding sequence can be

30

- 47 -

modified to code for the substitution, and the selectable marker can be placed outside of the coding region, for example, in a nearby intron.

The selectable marker may depend on its own promoter for expression and the marker  
5 gene may be derived from a very different organism than the organism being targeted (e.g. prokaryotic marker genes used in targeting mammalian cells). However, it is preferable to replace the original promoter with transcriptional machinery known to function in the recipient cells. A large number of transcriptional initiation regions are available for such purposes including, for example, metallothionein promoters, thymidine kinase promoters,  
10  $\beta$ -actin promoters, immunoglobulin promoters, SV40 promoters and human cytomegalovirus promoters. A widely used example is the pSV2-*neo* plasmid which has the bacterial neomycin phosphotransferase gene under control of the SV40 early promoter and confers in mammalian cells resistance to G418 (an antibiotic related to neomycin). A number of other variations may be employed to enhance expression of the selectable  
15 markers in animal cells, such as the addition of a poly(A) sequence and the addition of synthetic translation initiation sequences. Both constitutive and inducible promoters may be used.

The DNA is preferably modified by homologous recombination. The target DNA can be in  
20 any organelle of the animal cell including the nucleus and mitochondria and can be an intact gene, an exon or intron, a regulatory sequence or any region between genes.

Homologous DNA is a DNA sequence that is at least 70% identical with a reference DNA sequence. An indication that two sequences are homologous is that they will hybridize  
25 with each other under stringent conditions (Sambrook *et al.*, 1990, *supra*).

The genetically modified animals contemplated herein include "knock out" or "knock in" animals or genetic sequencing carrying one or more nucleotide additions, deletions, substitutions and/or insertions. They are useful in a range of applications including the  
30 development of medical assessment systems such as to monitor particle physiological conditions including genetic defects such as but not limited to spinabifida in humans. The

- 48 -

medical assessment systems are also useful as a model for wound healing and closure and for agents which modulate same.

The present invention further contemplates conditional genetically modified animals, such as those produced using recombination methods that are standard in the art. Bacteriophage P1 Cre recombinase and flp recombinase from yeast plasmids are two non-limiting examples of site-specific DNA recombinase enzymes that leave DNA at specific target sites (box P sites for Cre recombinase and *fit* sites for flp recombinase).

- 10 The present invention further contemplates co-suppression (i.e. sense suppression) and antisense suppression to down-regulate expression of *M-grh*. This would generally occur in a target test animal such as to generate a disease model:

In addition to providing a diagnostic capability as described above, the isolated nucleic acid molecules of the present invention may also provide a therapeutic capability by being used to correct or complement an abnormality detected in a subject. To deliver the appropriate sequence to a recipient cell or tissue of a subject, an isolated nucleic acid molecule of the present invention may be cloned into a suitable genetic construct such as a suitable vector.

20 Accordingly, a further aspect of the present invention contemplates a genetic construct comprising a nucleotide sequence encoding an *M-grh* selected from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or a variant thereof or a nucleotide sequence having at least 60% similarity to one or more of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 or a variant thereof or a nucleotide sequence capable of hybridizing to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 or SEQ ID NO:15 under low stringency conditions or a variant thereof or a complementary form thereof.

30 A "vector" is a polynucleotide molecule, preferably a DNA molecule derived, for example,

- 49 -

from a plasmid, bacteriophage, or plant virus, into which a polynucleotide can be inserted or cloned. A vector preferably contains one or more unique restriction sites and can be capable of autonomous replication in a defined host cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integrable with the genome of the defined host  
5 such that the cloned sequence is reproducible. Accordingly, the vector may be an autonomously replicating vector, i.e. a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication. Examples include a linear or closed circular plasmid, an extra-chromosomal element, a mini-chromosome, or an artificial chromosome. The vector may also contain a means for assuring self-  
10 replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. A vector system may comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon.

15

Vectors suitable for gene therapy applications are well known in the art. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which it is to be introduced. The vector may also include an additional genetic construct comprising a selection marker such as an antibiotic resistance gene that can be used for  
20 selection of suitable transformants. Examples of such resistance genes are known to those skilled in the art and include the *nptII* gene that confers resistance to the antibiotics kanamycin, and G418 (Geneticin®) and the *hph* gene which confer resistance to the antibiotic hygromycin B.

25 Accordingly, in a related embodiment, the present invention provides a genetic construct comprising a promoter or functional equivalent thereof operably linked to a nucleotide sequence of the invention.

Reference herein to a "promoter" is to be taken in its broadest context and includes the  
30 transcriptional regulatory sequences of a classical genomic gene, which is required for accurate transcription initiation, with or without a CCAAT box sequence and additional

- 50 -

regulatory elements (i.e. upstream activating sequences, enhancers and silencers), which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. A promoter is usually, but not necessarily, positioned upstream (5') of a gene region, the expression of which it regulates. Furthermore, the regulatory elements comprising a promoter are usually positioned within 2 kb of the start site of transcription of the gene. As is known in the art, some variation in this distance can be accommodated without loss of promoter function.

The selection of an appropriate promoter sequence to regulate expression of a transcription factor encoded by an isolated nucleic acid molecule of the present invention is an important consideration. Examples of suitable promoters include viral, fungal, bacterial, animal and plant derived promoters capable of functioning in eukaryotic animal cells and, especially, human cells. The promoter may regulate the expression of the nucleic acid molecule differentially with respect to the cell, tissue or organ in which expression occurs, or with respect to the developmental stage at which expression occurs.

Preferably, the promoter is capable of regulating expression of a nucleic acid molecule in a eukaryotic cell, tissue or organ, at least during the period of time over which the regulated gene is expressed therein, and more preferably also immediately preceding the commencement of detectable expression of the regulated gene in said cell, tissue or organ.

Particularly preferred promoters for use with the nucleic acid molecules of the present invention include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, *lac* operator-promoter, *tac* promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, CaMV 35S promoter, SCSV promoter, SCBV promoter and the like. Those skilled in the art will readily be aware of additional promoter sequences other than those specifically described.

In the present context, the terms "in operable connection with" or "operably linked" or similar shall be taken to indicate that expression of the nucleic acid molecule is under the control of the promoter sequence, with which it is spatially connected, in a cell, tissue,

- 51 -

organ or whole organism.

The genetic construct of the present invention may also comprise a 3' non-translated sequence. A 3' non-translated sequence refers to that portion of a gene comprising a DNA  
5 segment that contains a polyadenylation signal and any other regulatory signals capable of effecting mRNA processing or gene expression. The polyadenylation signal is characterized by effecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. Polyadenylation signals are commonly recognized by the presence of homology to the canonical form 5' AATAAA-3' although variations are not uncommon.

10 Accordingly, a genetic construct comprising a nucleic acid molecule of the present invention, operably linked to a promoter, may be cloned into a suitable vector for delivery to a cell or tissue in which regulation is faulty, malfunctioning or non-existent, in order to rectify and/or provide the appropriate regulation. Vectors comprising appropriate genetic  
15 constructs may be delivered into target eukaryotic cells by a number of different means well known to those skilled in the art of molecular biology.

The present invention further contemplates the use of an M-GRH or *M-grh* in the manufacture of a medicament for the treatment of a disease condition in a mammal such as  
20 a human.

The present invention is further directed to promoters and 3'- and 5'-regulatory regions associated with genomic forms of *M-grh* genes. These regions can be readily identified by, for example, chromosome walking using *M-grh* nucleic acid molecules or probes or  
25 primers therefrom.

A further aspect of the present invention relates to the use of the invention in relation to the treatment and/or prophylaxis of disease conditions. Without limiting the present invention to any one theory or mode of action, the broad range of cellular functional activities which  
30 are regulated by transcription factors renders the regulation of transcription factor function an integral component of every aspect of both healthy and disease state physiological

- 52 -

processes. Accordingly, the method of the present invention provides a valuable tool for modulating aberrant or otherwise unwanted cellular functional activity which is regulated via transcription factors.

5 Accordingly, another aspect of the present invention is directed to a method for the treatment and/or prophylaxis of a condition in a subject, which condition is characterised by aberrant, unwanted or otherwise inappropriate cellular activity, said method comprising administering to said mammal an effective amount of an agent for a time and under conditions sufficient to modulate transcription factor function.

10

The terms "agent", "compound", "active agent", "pharmacologically active agent", "medicament", "active" and "drug" are used interchangeably herein to refer to a chemical compound that induces a desired pharmacological and/or physiological effect. The terms also encompass pharmaceutically acceptable and pharmacologically active ingredients of those active agents specifically mentioned herein including but not limited to salts, esters, amides, prodrugs, active metabolites, analogs and the like. When the terms "agent", "compound", "active agent", "pharmacologically active agent", "medicament", "active" and "drug" are used, then it is to be understood that this includes the active agent *per se* as well as pharmaceutically acceptable, pharmacologically active salts, esters, amides, prodrugs, metabolites, analogs, etc. The term "compound" is not to be construed as a chemical compound only but extends to peptides, polypeptides and proteins as well as genetic molecules such as RNA, DNA and chemical analogs thereof as well as RNAi- or siRNA-type molecules or complexes comprising same. In accordance with the previous aspects of the present invention, the agent preferably comprises a transcription factor or genetic molecules encoding same or derivative, analogue, chemical equivalent or mimetic thereof.

15  
20  
25

"Subject" as used herein refers to an animal, preferably a mammal and more preferably human who can benefit from the pharmaceutical formulations and methods of the present invention. There is no limitation on the type of animal that could benefit from the presently described pharmaceutical formulations and methods. A patient regardless of whether a

30



- 53 -

human or non-human animal may be referred to as an individual, subject, animal, host or recipient.

The preferred animals are humans or other primates, livestock animals, laboratory test  
5 animals, companion animals or captive wild animals.

Examples of laboratory test animals include mice, rats, rabbits, guinea pigs and hamsters. Rabbits and rodent animals, such as rats and mice, provide a convenient test system or animal model. Livestock animals include sheep, cows, pigs, goats, horses and donkeys.  
10 Non-mammalian animals such as zebrafish and amphibians (including cane toads) are also contemplated

An "effective amount" means an amount necessary at least partly to attain the desired response, or to delay the onset or inhibit progression or halt altogether, the onset or  
15 progression of a particular condition being treated. The amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated, the degree of protection desired, the formulation of the composition, the assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined  
20 through routine trials.

Reference herein to "treatment" and "prophylaxis" is to be considered in its broadest context. The term "treatment" does not necessarily imply that a subject is treated until total recovery. Similarly, "prophylaxis" does not necessarily mean that the subject will not  
25 eventually contract a disease condition. Accordingly, treatment and prophylaxis include amelioration of the symptoms of a particular condition or preventing or otherwise reducing the risk of developing a particular condition. The term "prophylaxis" may be considered as reducing the severity or onset of a particular condition. "Treatment" may also reduce the severity of an existing condition.

30

- 54 -

The present invention further contemplates a combination of therapies, such as the administration of the agent together with subjection of the mammal to other agents, drugs or treatments which may be useful in relation to the treatment of the subject condition such as spina bifida and anencephaly.

5

Administration of the modulatory agent, in the form of a pharmaceutical composition, may be performed by any convenient means. The modulatory agent of the pharmaceutical composition is contemplated to exhibit therapeutic activity when administered in an amount which depends on the particular case. The variation depends, for example, on the  
10 human or animal and the modulatory agent chosen. A broad range of doses may be applicable. Considering a patient, for example, from about 0.1mg, 0.2mg, 0.3mg, 0.4mg, 0.5mg, 0.6mg, 0.7mg, 0.8mg, 0.9mg to about 1 mg of modulatory agent may be administered per kilogram of body weight per day. Dosage regimes may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be  
15 administered daily, weekly, monthly or other suitable time intervals or the dose may be proportionally reduced as indicated by the exigencies of the situation.

The modulatory agent may be administered in a convenient manner such as by the oral, intravenous (where water soluble), intraperitoneal, intramuscular, subcutaneous,  
20 intradermal or suppository routes or implanting (e.g. using slow release molecules). The modulatory agent may be administered in the form of pharmaceutically acceptable nontoxic salts, such as acid addition salts or metal complexes, e.g. with zinc, iron or the like (which are considered as salts for purposes of this application). Illustrative of such acid addition salts are hydrochloride, hydrobromide, sulphate, phosphate, maleate, acetate,  
25 citrate, benzoate, succinate, malate, ascorbate, tartrate and the like. If the active ingredient is to be administered in tablet form, the tablet may contain a binder such as tragacanth, corn starch or gelatin; a disintegrating agent, such as alginic acid; and a lubricant, such as magnesium stearate.

30 Routes of administration include, but are not limited to, respiratorally, intratracheally, nasopharyngeally, intravenously, intraperitoneally, subcutaneously, intracranially,

- 55 -

intradermally, intramuscularly, intraocularly, intrathecally, intracerebrally, intranasally, infusion, orally, rectally, *via* IV drip patch and implant.

In accordance with these methods, the agent defined in accordance with the present invention may be coadministered with one or more other compounds or molecules. By "coadministered" is meant simultaneous administration in the same formulation or in two different formulations via the same or different routes or sequential administration by the same or different routes. For example, the subject agent may be administered together with an agonistic agent in order to enhance its effects. By "sequential" administration is meant a time difference of from seconds, minutes, hours or days between the administration of the two types of molecules. These molecules may be administered in any order.

Another aspect of the present invention contemplates the use of an agent, as hereinbefore defined, in the manufacture of medicament for the treatment of a condition in a subject, which condition is characterised by aberrant, unwanted or otherwise inappropriate cellular activity, wherein said agent modulates transcription factor function.

In yet another further aspect, the present invention contemplates a pharmaceutical composition comprising the modulatory agent as hereinbefore defined together with one or more pharmaceutically acceptable carriers and/or diluents. These agents are referred to as the active ingredients.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion or may be in the form of a cream or other form suitable for topical application. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and

- 56 -

vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilisation. Generally, dispersions are prepared by incorporating the various sterilised active ingredient into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained. Preferred compositions or preparations according to the

- 57 -

present invention are prepared so that an oral dosage unit form contains between about 0.1 µg and 2000 mg of active compound.

The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter: a binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

Antisense polynucleotide sequences are another useful example of a therapeutic agent which can prevent or diminish the expression of the transcription factor genetic sequences, as will be appreciated by those skilled in the art. Polynucleotide vectors, for example, containing all or a portion of the *M-grh* sequences or other sequences from an *M-grh* region (particularly those flanking an *M-grh* gene locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with gene transcription and/or translation. Furthermore, co-suppression and mechanisms to induce RNAi (i.e. siRNA) may also be employed. Such techniques may be useful to inhibit genes which positively promote *M-grh* gene expression. Alternatively, antisense or sense molecules may be directly administered. In this latter embodiment, the antisense or sense molecules may be formulated in a composition and then administered by any number of means to target cells.

- 58 -

A variation on antisense and sense molecules involves the use of morpholinos, which are oligonucleotides composed of morpholine nucleotide derivatives and phosphorodiamidate linkages (for example, Summerton and Weller, linkages (for example, Summerton and Weller, *Antisense and Nucleic Acid Drug Development* 7: 187-195, 1997). Such compounds are injected into embryos and the effect of interference with mRNA is observed.

In one embodiment, the present invention employs compounds such as oligonucleotides and similar species for use in modulating the function or effect of nucleic acid molecules encoding an M-GRH transcription factor, i.e. the oligonucleotides induce transcriptional or post-transcriptional gene silencing. This is accomplished by providing oligonucleotides which specifically hybridize with one or more nucleic acid molecules encoding the transcription factor. As used herein, the terms "target nucleic acid" and "nucleic acid molecule encoding a transcription factor" have been used for convenience to encompass DNA encoding M-GRH, RNA (including pre-mRNA and mRNA or portions thereof) transcribed from such DNA, and also cDNA derived from such RNA. The hybridization of a compound of the subject invention with its target nucleic acid is generally referred to as "antisense". Consequently, the preferred mechanism believed to be included in the practice of some preferred embodiments of the invention is referred to herein as "antisense inhibition." Such antisense inhibition is typically based upon hydrogen bonding-based hybridization of oligonucleotide strands or segments such that at least one strand or segment is cleaved, degraded, or otherwise rendered inoperable. In this regard, it is presently preferred to target specific nucleic acid molecules and their functions for such antisense inhibition.

The functions of DNA to be interfered with can include replication and transcription. Replication and transcription, for example, can be from an endogenous cellular template, a vector, a plasmid construct or otherwise. The functions of RNA to be interfered with can include functions such as translocation of the RNA to a site of protein translation, translocation of the RNA to sites within the cell which are distant from the site of RNA synthesis, translation of protein from the RNA, splicing of the RNA to yield one or more

- 59 -

RNA species, and catalytic activity or complex formation involving the RNA which may be engaged in or facilitated by the RNA. One preferred result of such interference with target nucleic acid function is modulation of the expression of a *M-grh* gene. In the context of the present invention, "modulation" and "modulation of expression" mean either an  
5 increase (stimulation) or a decrease (inhibition) in the amount or levels of a nucleic acid molecule encoding the gene, e.g., DNA or RNA. Inhibition is often the preferred form of modulation of expression and mRNA is often a preferred target nucleic acid.

In the context of this invention, "hybridization" means the pairing of complementary  
10 strands of oligomeric compounds. In the present invention, the preferred mechanism of pairing involves hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleoside or nucleotide bases (nucleobases) of the strands of oligomeric compounds. For example, adenine and thymine are complementary nucleobases which pair through the formation of hydrogen bonds.  
15 Hybridization can occur under varying circumstances.

An antisense compound is specifically hybridizable when binding of the compound to the target nucleic acid interferes with the normal function of the target nucleic acid to cause a loss of activity, and there is a sufficient degree of complementarity to avoid non-specific  
20 binding of the antisense compound to non-target nucleic acid sequences under conditions in which specific binding is desired, i.e. under physiological conditions in the case of *in vivo* assays or therapeutic treatment, and under conditions in which assays are performed in the case of *in vitro* assays.

25 "Complementary" as used herein, refers to the capacity for precise pairing between two nucleobases of an oligomeric compound. For example, if a nucleobase at a certain position of an oligonucleotide (an oligomeric compound), is capable of hydrogen bonding with a nucleobase at a certain position of a target nucleic acid, said target nucleic acid being a DNA, RNA, or oligonucleotide molecule, then the position of hydrogen bonding between  
30 the oligonucleotide and the target nucleic acid is considered to be a complementary position. The oligonucleotide and the further DNA, RNA, or oligonucleotide molecule are

- 60 -

complementary to each other when a sufficient number of complementary positions in each molecule are occupied by nucleobases which can hydrogen bond with each other. Thus, "specifically hybridizable" and "complementary" are terms which are used to indicate a sufficient degree of precise pairing or complementarity over a sufficient number  
5 of nucleobases such that stable and specific binding occurs between the oligonucleotide and a target nucleic acid.

According to the present invention, compounds include antisense oligomeric compounds, antisense oligonucleotides, ribozymes, external guide sequence (EGS) oligonucleotides,  
10 alternate splicers, primers, probes, and other oligomeric compounds which hybridize to at least a portion of the target nucleic acid. As such, these compounds may be introduced in the form of single-stranded, double-stranded, circular or hairpin oligomeric compounds and may contain structural elements such as internal or terminal bulges or loops. Once introduced to a system, the compounds of the invention may elicit the action of one or  
15 more enzymes or structural proteins to effect modification of the target nucleic acid. One non-limiting example of such an enzyme is RNase H, a cellular endonuclease which cleaves the RNA strand of an RNA:DNA duplex. It is known in the art that single-stranded antisense compounds which are "DNA-like" elicit RNase H. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency  
20 of oligonucleotide-mediated inhibition of gene expression. Similar roles have been postulated for other ribonucleases such as those in the RNase III and ribonuclease L family of enzymes.

While the preferred form of antisense compound is a single-stranded antisense  
25 oligonucleotide, in many species the introduction of double-stranded structures, such as double-stranded RNA (dsRNA) molecules, has been shown to induce potent and specific antisense-mediated reduction of the function of a gene or its associated gene products. This phenomenon occurs in both plants and animals.

30 In the context of the subject invention, the term "oligomeric compound" refers to a polymer or oligomer comprising a plurality of monomeric units. In the context of this



- 61 -

invention, the term "oligonucleotide" refers to an oligomer or polymer of ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) or mimetics, chimeras, analogs and homologs thereof. This term includes oligonucleotides composed of naturally occurring nucleobases, sugars and covalent internucleoside (backbone) linkages as well as oligonucleotides having  
5 non-naturally occurring portions which function similarly. Such modified or substituted oligonucleotides are often preferred over native forms because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for a target nucleic acid and increased stability in the presence of nucleases.

10 While oligonucleotides are a preferred form of the compounds of this invention, the present invention comprehends other families of compounds as well, including but not limited to oligonucleotide analogs and mimetics such as those described herein.

The compounds in accordance with this invention preferably comprise from about 8 to  
15 about 80 nucleobases (i.e. from about 8 to about 80 linked nucleosides). One of ordinary skill in the art will appreciate that the invention embodies compounds of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, or 80 nucleobases  
20 in length.

As is known in the art, a nucleoside is a base-sugar combination. The base portion of the nucleoside is normally a heterocyclic base. The two most common classes of such heterocyclic bases are the purines and the pyrimidines. Nucleotides are nucleosides that  
25 further include a phosphate group covalently linked to the sugar portion of the nucleoside. For those nucleosides that include a pentofuranosyl sugar, the phosphate group can be linked to either the 2', 3' or 5' hydroxyl moiety of the sugar. In forming oligonucleotides, the phosphate groups covalently link adjacent nucleosides to one another to form a linear polymeric compound. In turn, the respective ends of this linear polymeric compound can  
30 be further joined to form a circular compound, however, linear compounds are generally preferred. In addition, linear compounds may have internal nucleobase complementarity

- 62 -

and may therefore fold in a manner as to produce a fully or partially double-stranded compound. Within oligonucleotides, the phosphate groups are commonly referred to as forming the internucleoside backbone of the oligonucleotide. The normal linkage or backbone of RNA and DNA is a 3' to 5' phosphodiester linkage.

5

Specific examples of preferred antisense compounds useful in this invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. As defined in this specification, oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus  
10 atom in the backbone. For the purposes of this specification, and as sometimes referenced in the art, modified oligonucleotides that do not have a phosphorus atom in their internucleoside backbone can also be considered to be oligonucleosides.

Preferred modified oligonucleotide backbones containing a phosphorus atom therein  
15 include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates,  
20 thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage. Preferred oligonucleotides having inverted polarity comprise a single 3' to 3' linkage at the 3'-most internucleotide linkage i.e. a single inverted nucleoside residue which may be abasic (the  
25 nucleobase is missing or has a hydroxyl group in place thereof). Various salts, mixed salts and free acid forms are also included.

Many of the preferred features described above are appropriate for sense nucleic acid molecules.

30

Another aspect of the present invention contemplates a method for the treatment or

- 63 -

prophylaxis of an animal, said method comprising exposing stem cells in said animal to one or more agents comprising a transcription factor, or a genetic molecule encoding a transcription factor or derivative, analogue, chemical equivalent or mimetic thereof which facilitates the proliferation and/or differentiation and/or self-renewal of stem cells to  
5 facilitate repair, replacement or augmentation of particular tissue.

In a related embodiment, the present invention provides a method for the treatment or prophylaxis of an animal, said method comprising exposing stem cells and mature cells or cells developmentally in between in said animal to one or more agents which facilitate the  
10 proliferation and/or differentiation and/or self-renewal of stem cells and mature cells or of cells developmentally in between to facilitate repair, replacement or augmentation of particular tissue.

As indicated above, the term "animal" includes a human amongst a range of other animals  
15 including avian species. The agent may comprise a single molecule or a combination of two or more molecules in a synergistic combination, admixture or cocktail. When in combination, the agents may be administered or used simultaneously or used sequentially such as seconds, minutes, hours, days or weeks apart. As indicated further below, the agent including therapeutic agent of this aspect of the present invention may also be a multi-part  
20 pharmaceutical pack or composition with instructions for use.

Reference to "exposing" to stem cells includes the situation where an agent is introduced into the body of the animal or the agent is contacted to an internal or external surface of, for example, skin or an organ or is otherwise administered to a surface or sub-surface or  
25 internal region. Alternatively, stem cells are removed from the animal's body and exposed to the agent *ex vivo* to facilitate differentiation and/or proliferation and/or self-renewal (either *ex vivo* or *in vivo*) and then the cells are returned to the same or different individual. According to this aspect of the present invention, although it is preferred to administer a therapeutic agent to an animal subject (e.g. a human), part of the therapeutic protocol may  
30 occur *ex vivo*. For example, proliferation may occur *in vivo* and differentiation may occur *in vivo*. Alternatively, part proliferation and part differentiation may occur *ex vivo* and

- 64 -

further facilitated *in vivo* by the administration of a therapeutic agent. Yet in a further alternative, proliferation occurs *ex vivo* and partial differentiation occurs *in vitro* but complete differentiation occurs *in vivo*. The terms "*ex vivo*" and "*in vitro*" are used interchangeably in this specification. In some circumstances, stem cells maintained *in vitro* may be used. Alternatively, *ex vivo* cells may first be genetically modified prior to re-introduction into a subject's body or a compatible counterpart.

Reference to "*in vitro*" or "*ex vivo*" means in tissue culture or in any situation outside the animal body. The term "*in vivo*" also means *in situ* and means treatment inside an animal body.

A mature cell in this context also includes a committed cell. This aspect of the present invention also extends to fetal cells such as ES cells or EG cells.

The entire repertoire of stem cells may be targeted by the therapeutic agent or one or more sub-populations may be induced to proliferate and/or differentiate. This is the difference between a generic agent or a specific agent. Cell sub-populations contemplated by the present invention include cells from the brain (e.g. adult neural stem cells, neurons, astrocytes), epidermis (e.g. keratinocyte stem cells, keratinocyte transient amplifying cells, keratinocyte post-mitotic differentiating cells, melanocyte stem cells, melanocytes), embryos (e.g. ES or EG cells), skin (e.g. foreskin fibroblasts), pancreas (e.g. pancreatic islet cells, pancreatic  $\beta$  cells), kidney (e.g. adult renal stem cells, embryonic renal epithelial stem cells, kidney epithelial cells), liver (e.g. hepatocytes, bile duct epithelial cells, embryonic endodermal stem cells, adult hepatocyte stem cells), breast (e.g. mammary epithelial stem cells), lung (e.g. bone marrow-derived stem cells, lung fibroblasts, bronchial epithelial cells, alveolar type II pneumocytes), muscle (e.g. skeletal muscle stem cells [satellite cells]), heart (e.g. cardiomyocytes, bone marrow mesenchymal stem cells), eye (e.g. limbal stem cells, corneal epithelial cells), bone (e.g. mesenchymal stem cells, osteoblasts [precursor of mesenchymal stem cells], peripheral blood mononuclear progenitor cells [hematopoietic stem cells], osteoclasts), spleen (e.g. splenocytes) and cells from the immune system (e.g.  $CD34^{+}$  stem cells,  $CD11c^{+}$  cells,  $CD11c^{-}$  cells,  $CD4^{+}$  T-

- 65 -

cells, CD8<sup>+</sup> T-cells, NK cells, monocytes, macrophages, dendritic cells and  $\beta$ -cells.

Whilst some of the above-listed cells are "mature" cells, they nevertheless may participate in a repair, regeneration or augmentation process by being selectively proliferated or used to "hone" in on particular tissue in need of treatment. Accordingly, the present invention is not to be interpreted as excluding the participation of mature cell types in the repair, regeneration and/or augmentation process as well as any other cell at a developmental stage between an ES cell and a mature cell.

Accordingly, another aspect of the present invention contemplates a method for tissue repair, regeneration and/or augmentation in an animal, said method comprising administering to said animal an agent or a combination of two or more agents, which agents promote or otherwise facilitate the proliferation and/or differentiation and/or self-renewal of a cell type selected from the listing comprising adult neural stem cells, neurons, astrocytes, keratinocyte stem cells, keratinocyte transient amplifying cells, keratinocyte post-mitotic differentiating cells, melanocyte stem cells, melanocytes, embryonic stem cells, embryonic germ cells, foreskin fibroblasts, pancreatic islet cells, pancreatic  $\beta$ -cells, adult renal stem cells, embryonic renal epithelial stem cells, kidney epithelial cells, hepatocytes, bile duct epithelial cells, embryonic endodermal stem cells, adult hepatocyte stem cells, mammary epithelial stem cells, bone marrow-derived stem cells, lung fibroblasts, bronchial epithelial cells, alveolar type II pneumocytes, skeletal muscle stem cells [satellite cells], cardiomyocytes, bone marrow mesenchymal stem cells, limbal stem cells, corneal epithelial cells, mesenchymal stem cells, osteoblasts [precursor of mesenchymal stem cells], peripheral blood mononuclear progenitor cells [hematopoietic stem cells], osteoclasts or splenocytes, said agents being administered for a time and under conditions sufficient to promote tissue repair, augmentation and/or regeneration.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in

- 66 -

this specification, individually or collectively, and any and all combinations of any two or more of said steps or features. It is also to be understood that unless stated otherwise, the subject invention is not limited to specific formulation components, manufacturing methods, dosage regimes, or the like, as such may vary.

5

The present invention is further described by the following non-limiting Examples.

### EXAMPLE 1

#### *Polymerase chain reaction*

10

For RT-PCR, first strand cDNA was prepared from 2 µg of mRNA from primary tissues using random hexamers. Each cDNA sample was appropriately diluted to give similar amplification of S14 RNA under the same PCR conditions. The primer sequences are detailed below. The PCR conditions were 94°C for 2 min followed by 35 cycles of 94°C for 30 sec, 60°C for 30 sec and 72°C for 45 sec with a final extension at 72°C for 5 min. All PCR products were electrophoresed on 1.5% w/v agarose gels, transferred to nitrocellulose and analyzed by Southern blot using <sup>32</sup>P-radiolabeled internal oligonucleotides as probes. Membranes were then autoradiographed for 2 hr at -70°C.

15

20 The following primers were used to amplify probes for cDNA library screening and for RT-PCR:-

#### *human p49 mgr*

5'-GAAGTCTTTGATGCCCTGATG-3' [SEQ ID NO:19]

25 5'-AACCCATTCCCTCGACATAGA-3' [SEQ ID NO:20]

#### *human p70 mgr*

5-AGCGCGATGACACAGGAGTA-3' [SEQ ID NO:21]

5'-CGTTGCTATGGAGACAGTGA-3' [SEQ ID NO:22]

30

- 67 -

*human bom*

5'-CCGTTTAACAAGGACACTGC-3' [SEQ ID NO:23]

5'-CTGGAAGCCACCAAATCTCT-3' [SEQ ID NO:24]

5 *murine p70 mgr*

5'-AGCGCGATGACACAGGAGTA-3' [SEQ ID NO:25]

5'-AGTGCCAGAGCTGAACTGAT-3' [SEQ ID NO:26]

*murine p61 mgr*

10 5'-TCCATGGGTTTCCTTGAGTTC-3' [SEQ ID NO:27]

5'-AGTGCCAGAGCTGAACTGAT'-3' [SEQ ID NO:28]

*murine bom*

5'-AAAGGGGAGCGAGTTCATTG-3' [SEQ ID NO:29]

15 5'-AGAGCTCTCGGTGATGGATA-3' [SEQ ID NO:30]

**EXAMPLE 2***Cloning of human and murine mgr and bom*

20

Human p49 *mgr* was cloned from a fetal brain cDNA phage library in the λZAP II vector (Stratagene). The cDNA encoding the longer human MGR isoform was amplified by RT-PCR from human kidney mRNA. The cDNA encoding the smaller murine isoform of MGR was cloned from a 17.5-day embryo phage library in the Lambda TripelEx vector

25 The murine p70 cDNA was amplified from murine kidney mRNA by RT-PCR. The human *bom* cDNA was isolated from a placental phage library in the Lambda ZAP II vector (Stratagene) and the murine cDNA from an embryonic carcinoma cell line (P19) phage library in the Uni-ZAP XR vector (Stratagene). The murine MGR genomic locus was obtained from a 129SVJ phage library in the Lambda FLX II vector (Stratagene).

30

- 68 -

From similarity searches of GenBank databases, using the GRH protein sequence as a query, two murine expressed sequence tag (EST) entries were found from adult brain and ovary and one human EST entry from fetal brain that were not identical to any previously reported genes, yet shared high degrees of homology with each other and *grh*. These sequences were used to design murine and human primers and amplified probes from murine adult brain and ovary and human adult brain cDNA. The murine probe from adult brain cDNA was used in a screen of a day 17.5 mouse embryo cDNA library to obtain a full length clone of a gene referred to as mammalian *grainyhead* (*mgr*) due to its sequence and functional homology and similar expression pattern to that of the fly gene. The human probe derived from adult brain cDNA was used to obtain a full length cDNA clone from a human fetal brain library. Amino acid sequence comparison reveals this to be the human homolog of MGR with 94% identity at the amino acid level.

The murine probe derived from ovary cDNA was used in a screen of a murine teratocarcinoma cell line (P19) cDNA library to obtain a full length clone of a novel gene distinct from but highly related to *mgr* named *brother-of-mgr* (*bom*). The homology between *mgr* and *bom* suggests that *mgr* and *bom* arose through gene duplication.

The human homolog of *bom* was obtained using primers derived from a high throughput genome sequencing (HTGS) database entry with homology to murine *bom*. These were used to amplify a probe from a human placental cDNA library that was then screened to yield a full length human cDNA clone. Amino acid sequence comparison between murine and human BOM revealed 94% identity.

The sequence alignments between *grh*, *mgr*, *bom*, *CP2* and *LBP-1a* revealed that *mgr* and *bom* are more closely related to *grh* than the previously identified homologs *CP2* and *LBP-1a* (Table 4). This homology is particularly evident in the DNA binding and dimerization domains emphasizing the importance of protein/protein and protein/DNA interactions for the function of these factors.



**TABLE 4** Amino acid sequence comparison of GRH-like genes and *Drosophila grh*

Amino acid identity/ similarity to Grainyhead (%)	Overall	DNA-binding domain	Dimerization domain
MGR	37/52	48/64	39/61
BOM	35/52	46/63	37/61
SOM	33/48	42/60	38/57
CP2	26/42	32/52	29/47
LBP-1a	23/39	31/51	28/43

**EXAMPLE 3***Identification of a second isoform of MGR*

A striking feature of the alignment between MGR and BOM was the absence of an MGR domain corresponding to the first 93 amino acids of BOM. In view of the absence of tissue-specific isoforms of GRH, the EST database was searched for similar sequences using the 5' end of *bom* as a query. A highly similar but non-identical sequence in an EST from murine kidney was located. The most 3' 30 nucleotides of this EST was identical to 30 nucleotides close to the 5' end of the *mgr*. Based on this, primers were designed from the kidney EST and *mgr* cDNA sequences and amplified a product of the predicted size from murine kidney cDNA. A similar product was also amplified from human kidney cDNA. Amino acid sequence analysis of the murine product revealed that it was highly homologous to the 5' end of the BOM protein and contiguous with the *mgr* open reading frame. However, it lacked the first 11 amino acids of a previously isolated *mgr* clone suggesting the presence of alternate splicing. To examine this, the murine *mgr* genomic locus was isolated and mapped. As shown in Figure 1B, the first three coding exons in the locus are exclusive to the p70 isoform of *mgr*. In contrast, the shorter isoform of *mgr*'s (p61) first coding exon is absent in the p70 isoform. Significantly, the 5' end of this exon lacks a splice acceptor site explaining its absence from the longer isoform. Instead, promoter sequences with a clear TATA box and CAP site are evident in close proximity to the translation initiation site (Figure 1C). Subsequent mapping of the human genomic locus revealed that murine exon four was conserved in the human p70 protein but was absent in

- 70 -

the 49 kDa isoform of MGR.

#### EXAMPLE 4

##### *The first three exons of the mgr genomic locus encode transcriptional activation domain*

Although significant sequence homology exists between *grh* and the shorter *mgr* isoforms and p70 *mgr*, the isoleucine rich transcriptional activation domain identified in the fly protein is not conserved. Examination of the MGR-coding sequences failed to reveal a region homologous to other known transactivation domains. In view of the high degree of conservation of the first three coding exons of p70 *mgr* and *bom*, it was postulated that this could be the functional domain responsible for activation. To address this, the cDNA fragment encoding the first 93 amino acids of human p70 MGR (encoded by the first three exons) was subcloned in frame with the GAL4 DNA binding domain in a mammalian expression vector. The comparable region of BOM and the full length p49 MGR cDNA in frame into this vector was also cloned. These plasmids were co-transfected into the human 293T cell line with a reporter plasmid containing five concatamerized GAL4 DNA binding sites upstream of the chloramphenicol acetyltransferase (CAT) gene. The vector containing only the GAL4 DNA-BD or containing the VP16 activation domain fused to the GAL DNA-BD served as the negative and positive controls, respectively. As shown in Figure 3, transcriptional activation of the CAT gene was observed with VP16, p70 MGR and the *bom* containing plasmids. No activation was observed with p49 *mgr* or the empty vector. These findings confirm the presence of a highly conserved activation domain in the p70 *mgr* and *bom* that is lacking in p49 *mgr*.

#### EXAMPLE 5

##### *MGR binds to known GRH binding sites*

To determine the extent of the functional homology between GRH and MGR, it was initially examined whether the mammalian protein could bind to the well-characterized binding sites for the *Drosophila* factor in the *Dopa decarboxylase* and *PCNA* gene

- 71 -

regulatory regions (Uv *et al.*, *Mol. Cell. Biol.* 17: 6727-6735, 1997; Hayashi *et al.*, *J. Biol. Chem.* 274: 35080-35088, 1999). Oligonucleotide probes encompassing these sites were incubated with nuclear extract from the human placental cell line JEG-3, which expresses both isoforms of MGR at RNA and protein level and analyzed in an electrophoretic mobility shift assay (EMSA).

EMSA were performed as previously described (Jane *et al.*, *EMBO J.* 14: 97-105, 1995) with the following oligonucleotide probes (sense strand only given): *Drosophila dopa decarboxylase* promoter (Uv *et al.*, 1997, *supra*) - GGTGGTGCTCTAATAACCGGTTT-CCAAGATGCGC (SEQ ID NO:31]; *Drosophila PCNA* promoter (Hayashi *et al.*, 1999, *supra*) - GGGTAAAAAGTGTGAACAATCAAACCAGTTGGCA (SEQ ID NO:32]; human Engrailed-1 promoter (Logan *et al.*, *Dev. Genet.* 13: 345-358, 1992) - GGACACACACCCAAACCCACACCCACCCACAAACACACAAACCGGCAGTGACAACAACCACCCATCCTTCAATAACAGCAACCA [SEQ ID NO:33]. In some assays, anti-MGR polyclonal antiserum was included in the reaction mix. Two antisera were used for this purpose: antisera 611 - raised against peptides common to the p70 and p49 MGR proteins in the dimerization domain; and antisera 67 raised against unique peptides in the NH<sub>2</sub>-terminal domain of p70 MGR. Nuclear extract for these assays was obtained from the human placental cell line, JEG-3.

As shown in Figure 2A, a specific protein/DNA complex was observed with the *PCNA* probe in the presence of pre-immune sera (lanes 1 and 3). This complex was supershifted with the addition of anti-p70 specific antisera raised against peptides in the amino terminal region of the protein (lane 4) and ablated with the addition of anti-MGR antisera raised against peptides common to p49 and p70 MGR in the dimerization domain of the protein (lane 2). Neither antisera cross-reacted with BOM. Similar results were obtained with the *Dopa decarboxylase* promoter probe (Figure 2B).

- 72 -

**EXAMPLE 6*****MGR binds to the human Engrailed-1 promoter***

Many *Drosophila* genes regulated by GRH have known mammalian homologs. In terms of functional homology, Engrailed-1 (En-1) is one of the bests characterized. The En-1 promoter was, therefore, examined for the *grainyhead* consensus DNA binding sequence derived from a comparison of the *Drosophila Ultrabithorax*, *Dopa decarboxylase* and *fushi tarazu* promoters (Dynlacht *et al.*, *Genes Dev.* 3: 1677-1688, 1989). As shown in Figure 3A, a highly conserved region was identified in the proximal En-1 promoter. Moreover, this sequence was also largely conserved in the DNaseI footprint attributed to *grh* in the *Drosophila engrailed* promoter (Soeller *et al.*, *Genes Dev.* 2: 68-81, 1988). The ability of this region of the human En-1 promoter to compete off MGR binding to the *Ddc* probe (Figure 3B) in an EMSA with nuclear extract from JEG-3 cells was examined. As shown in Figure 3B, the specific MGR/DNA complex observed with the *Ddc* probe (lane 1) was supershifted with the addition of MGR antisera 67 (lane 2) and ablated with the addition of a 50-fold excess of unlabeled *Ddc* probe as competitor (lane 3). Addition of a 10- (lane 4) or 20-fold (lane 5) excess of unlabeled En-1 probe also markedly reduced the binding of MGR to the *Ddc* probe.

**EXAMPLE 7*****MGR activates transcription***

To determine the functional significance of this binding, this region of the En-1 promoter was linked to a minimal globin gene promoter/luciferase reporter gene construct and transfected it into the MGR null cell line COS, in the presence of p70 MGR mammalian expression vector or the empty vector. Transfection of the minimal promoter/reporter or the TK promoter linked to a Renilla luciferase gene with either vector served as the controls. As shown in Figure 3C, expression of p70 MGR dramatically enhanced the transcriptional activity of the En-1 promoter (solid bars) but not the control minimal promoter (open bars) or the TK promoter (hatched bars).

- 73 -

**EXAMPLE 8*****Cloning of full-length human SOM***

Human SOM was cloned using primers derived from a high through-put genomic sequence  
5 (HTGS) and a human expression sequence tag (EST) obtained from GenBank databases  
which, respectively, aligned with the dimerization domain and the activation domain of  
other MGR members. Using nested RT-PCR and human tonsil cDNA, another contig  
spanning 1300 nucleotides was obtained.

10 Utilizing 5' RACE, further oligoprimers and human testis cDNA, a 210 nucleotide  
sequence incorporating the initiating ATG was obtained. A contig of these overlapping  
sequences revealed the full length human SOM which upon alignment with other existing  
MGR family members showed >60% similarity at the protein level with conservation at  
the 5' activation, DNA-binding and dimerization domains.

15

**EXAMPLE 9*****Cloning of full-length murine SOM***

A murine EST (GenBank) from optic cup tissue was identified, which when aligned with  
20 other murine homologs of the MGR family showed 70% similarity at the amino acid level,  
in the region of the DNA binding domain. Using semi-nested RT-PCR with murine testis  
cDNA, a 286 nucleotide sequence was amplified, cloned and sequenced for use as a probe.

Subsequently, a murine brain cDNA library (Stratagene) was screened. One clone was  
25 taken through to quaternary stage. This clone was excised from  $\lambda$ ZAP II vector into  
pBluescript and sequenced in both directions. A 1200 nucleotide length sequence was  
obtained, whichi lacked the 5' end. This was subsequently identified using 5' RACE from  
murine testis cDNA. A contig of these two sequences revealed the full length murine  
SOM.

30

- 74 -

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in  
5 this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

### EXAMPLE 10

#### *Generation of mice heterozygous at the GRHL-3 locus*

10

The murine GRHL-3 locus was isolated by screening a 129/SV/J genomic library with a cDNA fragment derived from the 5' end of the gene. Polymerase chain reaction (PCR) was then used to generate a 5.8 kb *NotI*-*Bam*HI fragment that when cloned into the plasmid p $\beta$ galpAloxneo, fused the second codon of GRHL-3 exon 2 to the ATG of  $\beta$ -galactosidase.  
15 The 3' flanking region was a *Sa*II-*Kpn*I fragment extending 2.6 kb from the beginning of intron 3. The thymidine kinase gene driven off the MC1-promoter was inserted into the targeting construct distal to the 3' arm as a *Sac*II-*Not*I fragment and the vector was linearized with *Not*I and electroporated into W9.5 embryonic stem cells. Transfected cells were selected in G418 and resistant clones picked and expanded. Clones were identified in  
20 which the targeting vector had recombined with the endogenous GRHL-3 gene by hybridising *Spe*I digested genomic DNA with a 0.5 kb *Spe*I-*Sa*II fragment situated in the 5' GRHL-3 genomic sequence just outside the targeting vector. This probe distinguished between the endogenous (8.4 kb) and targeted (11.5 kb) GRHL-3 alleles. Two correctly targeted embryonic stem cell clones were injected into C57BL/6 blastocysts to generate  
25 chimeric mice. Male chimeras were mated with C57BL/6 females to yield GRHL-3 heterozygotes which were identified by hybridising *Bam*HI-digested genomic DNA from a tail biopsy with a 0.85 kb *Nco*I fragment situated in the 3' GRHL-3 genomic sequence just outside the targeting vector. This probe distinguished between the endogenous (5.2 kb) and targeted (10.7 kb) alleles. Heterozygous mice were bred with Cre deleter transgenic mice  
30 to excise the Neo<sup>R</sup> cassette. GRHL-3 heterozygotes in which the Neo<sup>R</sup> cassette had been deleted were interbred to produce wild type (GRHL-3<sup>+/+</sup>), heterozygous (GRHL-3<sup>+/-</sup>) and

- 75 -

mutant (GRHL-3<sup>-/-</sup>) mice. The inability of the targeted allele to produce GRHL-3 messenger RNA was confirmed in nucleic acid blots.

C57BL/6J inbred mice were obtained from the Walter and Eliza Hall Institute animal facility, and the *ct/ct* mouse stock from the Jackson Laboratory. All experiments were  
5 approved by the Melbourne Health Animal Ethics Committee.

### EXAMPLE 11

#### *Genotyping GRHL-3 mutant mice*

10 Mice were genotyped by PCR using genomic DNA template prepared from tail biopsies or embryonic tissues. Products of 812 bp were generated from the wild type GRHL-3 allele and/or a product of 579 bp was generated from the targeted GRHL-3 allele. Primers used were specific for intron 1, common to the wild type and targeted GRHL-3 alleles (sense, 5'-GGATCAGAAGACCATGCC-3') (SEQ ID NO:40); intron 2, deleted from the targeted  
15 GRHL-3 allele (antisense, 5'-AGGCTGTAGAGTTGGTG-3') (SEQ ID NO:41); and the lacZ cassette, present only in the targeted GRHL-3 allele (antisense, 5'-CTGTAGCCAGCTTTCATC-3') (SEQ ID NO:42). PCR conditions were 94 °C for 2 minutes followed by 35 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 1 minute with a final 5 minutes extension at 72 °C.

20

### EXAMPLE 12

#### *Inositol and folate administration during pregnancy*

GRHL-3<sup>+/-</sup> mice were inter-crossed and folate, inositol or PBS placebo was administered to  
25 pregnant females as previously described. Embryos were harvested on E14.5 and genotyped and examined morphologically. Mean litter size and frequency of resorptions did not differ significantly for folate-, inositol- or placebo-treated litters. The results were analysed statistically by the one-sided binomial probability test.

- 76 -

**EXAMPLE 13*****GRHL-3 Northern hybridization and RT-PCR***

A unique GRHL-3 cDNA probe from nucleotides 404 to 889 was hybridized to a blot  
5 containing 35 µg of total RNA from E14.5 *ct/ct* embryos. RT-PCR was performed on  
cDNA from DNaseI-treated total RNA (2 µg) isolated from whole E9.5 embryos (Rneasy,  
Qiagen) using a First Strand Synthesis for RT-PCR kit (Amersham). One-tenth of the total  
cDNA was used as the template in PCR reactions containing primers specific for HPRT  
(sense, 5'-GCTGGTGAAAAGGACCTCT-3' (SEQ ID NO:43); antisense, 5'-  
10 CACAGGACTAGAACACCTGC-3') (SEQ ID NO:44). The cDNA sample was then  
diluted to give similar amplification of HPRT under the same PCR conditions prior to use  
in PCR reactions containing GRHL-3-specific primers. E9.5 GRHL-3<sup>+/+</sup> and GRHL-3<sup>-/-</sup>  
embryo cDNA was amplified with specific primers annealing to exon 8 and exon 13  
(sense; 5'-CACATTGAAGAGGTGGC-3' (SEQ ID NO:45); antisense, 5'-  
15 AAGGGTGAGCAGGTTCGCTT-3') (SEQ ID NO:46). PCR conditions were 94 °C for 2  
minutes followed by various cycles of 94 °C for 30 seconds, 60 °C for 30 seconds and 72  
°C for 1 minute. All PCR products were electrophoresed on 1.5% agarose gels, transferred  
to nitrocellulose membranes and analyzed by Southern blotting using P-labelled internal  
oligonucleotides as probes.

20

Quantitative Real-Time RT-PCR was performed in a Rotorgene 2000 (Corbett Research,  
Australia) in a final volume of 20 µl. Reaction mixtures comprised 1X reaction buffer plus  
2.5 mM (HPRT) or 3 mM (GRHL-3) MgCl<sub>2</sub>, 0.05 mM dNTPs (Roche), 0.1µM gene-  
specific primers, 1U Taq (Fisher Biotech, Australia), a 1/10000 dilution of SYBR Green I  
25 (Molecular Probes, USA), and 2 µl of sample or standard. Cycling conditions were 94 °C  
for 15 seconds, 55 °C (GRHL-3) or 52 °C (HPRT) for 30 seconds and 72 °C for 30  
seconds. For each reaction, standard curves were generated and relative quantities of each  
transcript were calculated from this. The ratio of GRHL-3/HPRT normalised to wild type



- 77 -

E14.5 is shown. Error bars show the sum of the standard deviations for each sample as a proportion of the normalised signal. The identity of the PCR products was confirmed by melt curve analysis and agarose gel electrophoresis on a 1.5% agarose gel.

5

#### EXAMPLE 14

##### *Histology, in situ hybridization and whole mount skeletal staining of GRHL-3 mutant mice*

Embryos from timed pregnant C57BL/6J females and GRHL-3<sup>+/-</sup> intercrosses were  
10 immersion fixed in 4% paraformaldehyde in phosphate buffered saline, pH 7.3. The  
embryos were then embedded in paraffin wax before 8 µm sections were cut on a  
microtome and placed on gelatine-coated slides. For histological analysis, sections were  
stained with hematoxylin and eosin. *In situ* hybridisation was performed as described  
previously. A radio-labelled GRHL-3 antisense RNA probe was transcribed from a  
15 pBluescript II SK plasmid (Stratagene) carrying a 485 bp fragment of the GRHL-3 coding  
region (nt 404 to 889) using T7 RNA polymerase. All sections for *in situ* analysis were  
counter-stained with hematoxylin. *In situ* hybridisation was also performed using a full-  
length GRHL-3 probe and the same expression pattern was observed (data not shown).  
Hybridisation signal was similar to background levels in embryos homozygous for the  
20 GRHL-3 mutation. Whole mount skeletal staining on E17.5 embryos was performed as  
described previously.

#### EXAMPLE 15

##### *Expression of GRHL-3 during mouse development*

25

One criterion for defining specific neurulation genes is that they are expressed in the  
region of the folding neural plate at the appropriate developmental time point. To  
determine the pattern of expression of GRHL-3 during murine development, *in situ*  
hybridisation studies were performed using a probe specific for GRHL-3 (Fig. 4). In  
30 embryos at E8, it was observed that expression was confined to the non-neuronal ectoderm  
immediately adjacent to the neural plate that was undergoing folding to form the neural

- 78 -

tube (Fig. 4A,B). At later time-points, more widespread expression was observed in the surface ectoderm with a progressive increase from E12.5 to E15.5 (Fig. 4C,D). The pattern of expression at later time points is similar to the expression profiles of murine GRHL-1 and -2 (Ref. 8).

5

## EXAMPLE 16

### *NTDs in GRHL-3 mutant mice*

To determine the functional role of GRHL-3 during mouse development, a 2.2 kb deletion  
10 in GRHL-3 was generated by gene targeting (Fig. 5A-D). Northern blot and RT-PCR  
analysis indicated that the targeted GRHL-3 allele represented a null mutation (Fig. 5E,F).  
Genotyping of offspring from GRHL-3<sup>+/-</sup> intercrosses from mid and late gestation, showed  
that GRHL-3<sup>-/-</sup> mice were represented in Mendelian proportions up to E18.5. Of 874  
embryos examined on, or before this time, 191 (22%) were genotyped as GRHL-3<sup>-/-</sup>. No  
15 GRHL-3<sup>-/-</sup> embryos survived to weaning.

All GRHL-3<sup>-/-</sup> pups, without exception, displayed neural tube defects (NTDs). GRHL-3<sup>+/-</sup>  
mice were indistinguishable from their wild-type littermates. As most of the newborn  
GRHL-3<sup>-/-</sup> pups were cannibalised by their mothers, we examined the phenotype in more  
20 detail in developing embryos (Fig. 6A). All GRHL-3<sup>-/-</sup> embryos exhibited thoraco-lumbo-  
sacral spina bifida and curled tail and 3% had co-incident exencephaly. GRHL-3<sup>-/-</sup>  
embryos were also smaller than their littermate controls. Full body skeletal preparations  
demonstrated abnormalities in the vertebral column with kyphosis, splayed spinal  
processes and lack of vertebral arch formation (Fig. 6B,C). Transverse sections through the  
25 thoracic, lumbar and sacral regions at different developmental time points showed that  
spina bifida in GRHL-3<sup>-/-</sup> embryos was due to a primary failure of neural tube closure. The  
neural plate appeared to furrow normally with the formation of the median hinge point, but  
neural fold elevation did not occur and the neuro-epithelium remained convex throughout  
gestation (Fig. 6D).

30

## EXAMPLE 17

*The GRHL-3 gene is allelic with the ct gene*

Analysis of GRHL-3 mutant mice revealed phenotypic similarities between the GRHL-3<sup>-/-</sup> embryos and those reported in the *curly tail* mouse mutant<sup>4</sup>. The *curly tail* gene is incompletely penetrant, with homozygotes developing lumbo-sacral spina bifida aperta (12%), curled tail (50%) and exencephaly (3%). The *ct* locus had been mapped to murine chromosome 4, at position 63.4, close to the D4Mit69 marker. The NCBI STS database to ascertain the chromosomal localisation of GRHL-3 in mice. The mGRHL-3 gene is also located on chromosome 4, approximately 3kb from D4Mit157 at position 63.4. This marker, which lies within 800 kb of D4Mit69, had not been included in the original *curly tail* mapping studies. Both markers were subsequently identified in a recently deposited 13 Mb contiguous sequence from chromosome 4 (Fig. 7A). Also contained in this sequence were several genes that have previously been studied (and excluded) as *ct* candidate genes. GRHL-3 was positioned closer to the D4Mit69 marker than all of these excluded candidates. Genetic complementation studies were therefore performed using mice heterozygous for the null GRHL-3 allele and *ct/ct* mice. Embryos from these matings were harvested between E11.5 and E18.5 days and genotyped and examined morphologically. Among the 101 embryos obtained, NTDs were the only gross abnormalities observed (Fig. 7B). These were confined, without exception, to embryos with a GRHL-3<sup>+/-</sup>/*ct* genotype (Table 5). The incidence of spina bifida in mice carrying both mutant alleles was higher than reported for *curly tail* homozygotes (31% versus 12%), but the extent of the defect more closely resembled that of the *ct/ct* mice (lumbo-sacral spina bifida) than the GRHL-3<sup>-/-</sup> mice (thoraco-lumbo-sacral spina bifida). Tail flexion defects alone were identified in an additional 23% of embryos, all of which were genotyped as GRHL-3<sup>+/-</sup>/*ct*. Ten embryos carrying both mutant alleles appeared morphologically normal and the remaining 37% of embryos were unremarkable and genotyped as GRHL-3<sup>+/-</sup>/*ct*.

The expression of GRHL-3 in *curly tail* homozygous embryos was compared with wild type and GRHL-3<sup>+/-</sup> controls by Northern blotting (Fig. 7C). Densitometry of the GRHL-3 signal normalised to the 28S signal obtained by Phosphorimager analysis revealed a

- 80 -

significant reduction in the level of GRHL-3 mRNA in *ct/ct* embryos (19-34%) compared with wild type or GRHL-3 heterozygous controls. Real-time quantitative RT-PCR was performed on mRNA from these embryos and confirmed that the level of GRHL-3 expression in *ct/ct* embryos was reduced approximately 3-fold compared to wild type controls (Fig. 7D).

TABLE 5 Phenotypes of embryos from GRHL-3<sup>+/-</sup> X *ct/ct* crosses

Phenotype	Number	Genotype
Spina bifida + curly tail	31	GRHL-3 <sup>+/-</sup> / <i>ct</i> -31
Curled tail	23	GRHL-3 <sup>+/-</sup> / <i>ct</i> - 23
Normal	47	GRHL-3 <sup>+/-</sup> / <i>ct</i> - 10 GRHL-3 <sup>+/+</sup> / <i>ct</i> - 37
Total	101	101

Embryos were harvested between E11.5 and E18.5.

10

### EXAMPLE 18

#### *NTDs in GRHL-3<sup>-/-</sup> embryos are folate- and inositol-resistant*

15 NTDs in the *curly tail* mice are resistant to folate administered in early gestation. However, inositol therapy in pregnancy results in a marked reduction in the incidence of spina bifida. The effects of folate and inositol administration on pregnant GRHL-3<sup>+/-</sup> mice previously mated with GRHL-3<sup>+/-</sup> males was examined (Table 6). As expected, no rescue of spina bifida in GRHL-3<sup>-/-</sup> embryos with placebo or folate treatment was observed. In

- 81 -

contrast to the *ct/ct* mice, neither the incidence, nor severity of the spina bifida in the GRHL-3<sup>-/-</sup> embryos was alleviated by inositol treatment. Although the numbers of embryos examined was small, the result was highly significant ( $p < 0.001$ ) given the 70% reduction in the incidence of spina bifida in inositol-treated *ct/ct* embryos. These findings indicate

5 that GRHL-3 expression is essential for inositol-mediated rescue of folate-resistant NTDs.

**TABLE 6** Effects of folate and inositol administration on NTDs in GRHL-3<sup>-/-</sup> mice

Genotype GRHL-3 <sup>-/-</sup>	Placebo n=5	Folate n=4	Inositol n=8
Predicted NTDs	100%	100%	30%
Observed NTDs	100%	100%	100%*

\* $p < 0.001$